



# Estimating FARIMA models with uncorrelated but non-independent error terms

Yacouba Boubacar Maïnassara, Youssef Esstafa, Bruno Saussereau

## ► To cite this version:

Yacouba Boubacar Maïnassara, Youssef Esstafa, Bruno Saussereau. Estimating FARIMA models with uncorrelated but non-independent error terms. 2019. hal-02316503v1

**HAL Id: hal-02316503**

**<https://hal.science/hal-02316503v1>**

Preprint submitted on 15 Oct 2019 (v1), last revised 10 Mar 2021 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Estimating FARIMA models with uncorrelated but non-independent error terms

Yacouba Boubacar Maïnassara  
Youssef Esstafa  
Bruno Saussereau

*Université Bourgogne Franche-Comté,  
Laboratoire de mathématiques de Besançon,  
UMR CNRS 6623,  
16 route de Gray,  
25030 Besançon, France.*

*e-mail:* [yacouba.boubacar\\_mainassara@univ-fcomte.fr](mailto:yacouba.boubacar_mainassara@univ-fcomte.fr)  
*e-mail:* [youssef.esstafa@univ-fcomte.fr](mailto:youssef.esstafa@univ-fcomte.fr)  
*e-mail:* [bruno.saussereau@univ-fcomte.fr](mailto:bruno.saussereau@univ-fcomte.fr)

**Abstract:** In this paper we derive the asymptotic properties of the least squares estimator (LSE) of fractionally integrated autoregressive moving-average (FARIMA) models under the assumption that the errors are uncorrelated but not necessarily independent nor martingale differences. We relax considerably the independence and even the martingale difference assumptions on the innovation process to extend the range of application of the FARIMA models. We propose a consistent estimator of the asymptotic covariance matrix of the LSE which may be very different from that obtained in the standard framework. A self-normalized approach to confidence interval construction for weak FARIMA model parameters is also presented. All our results are done under a mixing assumption on the noise. Finally, some simulation studies and an application to the daily returns of stock market indices are presented to corroborate our theoretical work.

**AMS 2000 subject classifications:** Primary 62M10; secondary 91B84.

**Keywords and phrases:** Nonlinear processes; FARIMA models; Least-squares estimator; Consistency; Asymptotic normality; Spectral density estimation; Self-normalization; Cumulants.

## 1. Introduction

Long memory processes takes a large part in the literature of time series (see for instance [GJ80], [FT86], [Dah89], [Hos81], [BFGK13], [Pal07], among others). They also play an important role in many scientific disciplines and applied fields such as hydrology, climatology, economics, finance, to name a few. To model the long memory phenomenon, a widely used model is the fractional autoregressive integrated moving average (FARIMA, for short) model. Consider a second order centered stationary process  $X := (X_t)_{t \in \mathbb{Z}}$  satisfying a FARIMA( $p, d_0, q$ ) representation of the form

$$a(L)(1 - L)^{d_0}X_t = b(L)\epsilon_t, \quad (1)$$

where  $d_0 \in ]0, 1/2[$  is the long memory parameter,  $L$  stands for the back-shift operator and  $a(L) = 1 - \sum_{i=1}^p a_i L^i$  is the autoregressive (AR for short) operator and  $b(L) = 1 - \sum_{i=1}^q b_i L^i$  is the moving average (MA for short) operator (by convention  $a_0 = b_0 = 1$ ). The operators  $a$  and

$b$  represent the short memory part of the model. The linear innovation process  $\epsilon := (\epsilon_t)_{t \in \mathbb{Z}}$  is assumed to be a stationary sequence satisfies

(A0):  $\mathbb{E}[\epsilon_t] = 0$ ,  $\text{Var}(\epsilon_t) = \sigma_\epsilon^2$  and  $\text{Cov}(\epsilon_t, \epsilon_{t+h}) = 0$  for all  $t \in \mathbb{Z}$  and all  $h \neq 0$ .

Under the above assumptions the process  $\epsilon$  is called a weak white noise. Different sub-classes of FARIMA models can be distinguished depending on the noise assumptions. It is customary to say that  $X$  is a strong FARIMA( $p, d_0, q$ ) representation and we will do this henceforth if in (1)  $\epsilon$  is a strong white noise, namely an independent and identically distributed (iid for short) sequence of random variables with mean 0 and common variance. A strong white noise is obviously a weak white noise because independence entails uncorrelatedness. Of course the converse is not true. Between weak and strong noises, one can say that  $\epsilon$  is a semi-strong white noise if  $\epsilon$  is a stationary martingale difference, namely a sequence such that  $\mathbb{E}(\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \dots) = 0$ . An example of semi-strong white noise is the generalized autoregressive conditional heteroscedastic (GARCH) model (see [FZ10]). If  $\epsilon$  is a semi-strong white noise in (1),  $X$  is called a semi-strong FARIMA( $p, d_0, q$ ). If no additional assumption is made on  $\epsilon$ , that is if  $\epsilon$  is only a weak white noise (not necessarily iid, nor a martingale difference), the representation (1) is called a weak FARIMA( $p, d_0, q$ ). It is clear from these definitions that the following inclusions hold:

$$\{\text{strong FARIMA}(p, d_0, q)\} \subset \{\text{semi-strong FARIMA}(p, d_0, q)\} \subset \{\text{weak FARIMA}(p, d_0, q)\}.$$

Nonlinear models are becoming more and more employed because numerous real time series exhibit nonlinear dynamics. For instance conditional heteroscedasticity can not be generated by FARIMA models with iid noises.<sup>1</sup> As mentioned by [FZ05, FZ98] in the case of ARMA models, many important classes of nonlinear processes admit weak ARMA representations in which the linear innovation is not a martingale difference. The main issue with nonlinear models is that they are generally hard to identify and implement. These technical difficulties certainly explain the reason why the asymptotic theory of FARIMA model estimation is mainly limited to the strong or semi-strong FARIMA model.

Now we present some of the main works about FARIMA model estimation when the noise is strong or semi-strong. For the estimation of long-range dependent process, the commonly used estimation method is based on the Whittle frequency domain maximum likelihood estimator (MLE) (see for instance [Dah89], [FT86], [TT97], [GS90]). The asymptotic properties of the MLE of FARIMA models are well-known under the restrictive assumption that the errors  $\epsilon_t$  are independent or martingale difference (see [Ber95], [BFGK13], [Pal07], [BCT96], [LL97], [HK98], among others). All the works mentioned above assume either strong or semi-strong innovations. In the modeling of financial time series, for example, the GARCH assumption on the errors is often used (see for instance [BCT96], [HK98]) to capture the conditional heteroscedasticity. There is no doubt that it is important to have a soundness inference procedure for the parameter in the FARIMA model when the (possibly dependent) error is subject to unknown conditional heteroscedasticity. Little is thus known when the martingale difference assumption is relaxed. Our aim in this paper is to consider a flexible FARIMA specification and to relax the independence assumption (and even the martingale difference assumption) in order to be able to cover weak FARIMA representations of general nonlinear models. This is why it is interesting to consider weak FARIMA models.

---

<sup>1</sup> To cite few examples of nonlinear processes, let us mention the self-exciting threshold autoregressive (SETAR), the smooth transition autoregressive (STAR), the exponential autoregressive (EXPAR), the bilinear, the random coefficient autoregressive (RCA), the functional autoregressive (FAR) (see [Ton90] and [FY08] for references on these nonlinear time series models).

A very few works deal with the asymptotic behavior of the MLE of weak FARIMA models. To our knowledge, [Sha12, Sha10b] are the only papers on this subject. Under weak assumptions on the noise process, the author has obtained the asymptotic normality of the Whittle estimator (see [Whi53]). Nevertheless, the inference problem is not addressed. This is due to the fact that the asymptotic covariance matrix of the Whittle estimator involves the integral of the fourth-order cumulant spectra of the dependent errors  $\epsilon_t$ . Using non-parametric bandwidth-dependent methods, one build an estimation of this integral but there is no guidance on the choice of the bandwidth in the estimation procedures (see [Sha12, Tan82, Kee87, Chi88] for further details). The difficulty is caused by the dependence in  $\epsilon_t$ . Indeed, for strong noise, a bandwidth-free consistent estimator of the asymptotic covariance matrix is available. When  $\epsilon_t$  is dependent, no explicit formula for a consistent estimator of the asymptotic variance matrix seems to be provided in the literature (see [Sha12]).

In this work we propose to adopt for weak FARIMA models the estimation procedure developed in [FZ98] so we use the least squares estimator (LSE for short). We show that a strongly mixing property and the existence of moments are sufficient to obtain a consistent and asymptotically normally distributed least squares estimator for the parameters of a weak FARIMA representation. For technical reasons, we often use an assumption on the summability of cumulants. This can be a consequence of a mixing and moments assumptions (see [DL89], for more details). These kind of hypotheses enable us to circumvent the problem of the lack of speed of convergence (due to the long-range dependence) in the infinite AR or MA representations. We fix this gap by proposing rather sharp estimations of the infinite AR and MA representations in the presence of long-range dependence (see Subsection 6.1 for details).

In our opinion there are three major contributions in this work. The first one is to show that the estimation procedure developed in [FZ98] can be extended to weak FARIMA models. This goal is achieved thanks to Theorem 1 and Theorem 2 in which the consistency and the asymptotic normality are stated. The second one is to provide an answer to the open problem raised by [Sha12] (see also [Sha10b]) on the asymptotic covariance matrix estimation. We propose in our work a weakly consistent estimator of the asymptotic variance matrix (see Theorem 3). Thanks to this estimation of the asymptotic variance matrix, we can construct a confidence region for the estimation of the parameters. Finally another method to construct such confidence region is achieved thanks to an alternative method using a self normalization procedure (see Theorem 6).

The paper is organized as follows. Section 2 shown that the least squares estimator for the parameters of a weak FARIMA model is strongly consistent when the weak white noise ( $\epsilon_t$ ) is ergodic and stationary, and that the LSE is asymptotically normally distributed when ( $\epsilon_t$ ) satisfies mixing assumptions. The asymptotic variance of the LSE may be very different in the weak and strong cases. Section 3 is devoted to the estimation of this covariance matrix. We also propose a self-normalization-based approach to constructing a confidence region for the parameters of weak FARIMA models which avoids to estimate the asymptotic covariance matrix. We gather in Section 7 all our figures and tables. These simulation studies and illustrative applications on real data are presented and discussed in Section 4. The proofs of the main results are collected in Section 6.

In all this work, we shall use the matrix norm defined by  $\|A\| = \sup_{\|x\| \leq 1} \|Ax\| = \rho^{1/2}(A'A)$ , when  $A$  is a  $\mathbb{R}^{k_1 \times k_2}$  matrix,  $\|x\|^2 = x'x$  is the Euclidean norm of the vector  $x \in \mathbb{R}^{k_2}$ , and  $\rho(\cdot)$  denotes the spectral radius.

## 2. Least squares estimation

In this section we present the parametrization and the assumptions that are used in the sequel. Then we state the asymptotic properties of the LSE of weak FARIMA models.

### 2.1. Notations and assumptions

We make the following standard assumption on the roots of the AR and MA polynomials in (1).

**(A1):** The polynomials  $a(z)$  and  $b(z)$  have all their roots outside of the unit disk with no common factors.

Let  $\Theta^*$  be the compact space

$$\Theta^* := \left\{ (\theta_1, \theta_2, \dots, \theta_{p+q}) \in \mathbb{R}^{p+q}, \text{ where } a_\theta(z) = 1 - \sum_{i=1}^p \theta_i z^i, \text{ and } b_\theta(z) = 1 - \sum_{j=1}^q \theta_{p+j} z^j \right. \\ \left. \text{have all their zeros outside the unit disk and have no zero in common} \right\}.$$

Denote by  $\Theta$  the cartesian product  $\Theta^* \times [d_1, d_2]$ , where  $[d_1, d_2] \subset ]0, 1/2[$  with  $d_1 \leq d_0 \leq d_2$ . The unknown parameter of interest  $\theta_0 = (a_1, a_2, \dots, a_p, b_1, b_2, \dots, b_q, d_0)$  is supposed to belong to the parameter space  $\Theta$ .

The fractional difference operator  $(1 - L)^{d_0}$  is defined, using the generalized binomial series, by

$$(1 - L)^{d_0} = \sum_{j \geq 0} \alpha_j(d_0) L^j,$$

where for all  $j \geq 0$ ,  $\alpha_j(d_0) = \Gamma(j - d_0) / \{\Gamma(j + 1)\Gamma(-d_0)\}$  and  $\Gamma(\cdot)$  is the Gamma function. Using the Stirling formula we obtain that for large  $j$ ,  $\alpha_j(d_0) \sim j^{-d_0-1}/\Gamma(-d_0)$  (one refers to [BFGK13] for further details).

For all  $\theta \in \Theta$  we define  $(\epsilon_t(\theta))_{t \in \mathbb{Z}}$  as the second order stationary process which is the solution of

$$\epsilon_t(\theta) = \sum_{j \geq 0} \alpha_j(d) X_{t-j} - \sum_{i=1}^p \theta_i \sum_{j \geq 0} \alpha_j(d) X_{t-i-j} + \sum_{j=1}^q \theta_{p+j} \epsilon_{t-j}(\theta). \quad (2)$$

Observe that, for all  $t \in \mathbb{Z}$ ,  $\epsilon_t(\theta_0) = \epsilon_t$  a.s. Given a realization  $X_1, \dots, X_n$  of length  $n$ ,  $\epsilon_t(\theta)$  can be approximated, for  $0 < t \leq n$ , by  $\tilde{\epsilon}_t(\theta)$  defined recursively by

$$\tilde{\epsilon}_t(\theta) = \sum_{j=0}^{t-1} \alpha_j(d) X_{t-j} - \sum_{i=1}^p \theta_i \sum_{j=0}^{t-i-1} \alpha_j(d) X_{t-i-j} + \sum_{j=1}^q \theta_{p+j} \tilde{\epsilon}_{t-j}(\theta), \quad (3)$$

with  $\tilde{\epsilon}_t(\theta) = X_t = 0$  if  $t \leq 0$ . It will be shown that these initial values are asymptotically negligible uniformly in  $\theta$  and, in particular, that  $\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta) \rightarrow 0$  almost surely as  $t \rightarrow \infty$  (see Lemma 4 hereafter). Thus the choice of the initial values has no influence on the asymptotic properties of the model parameters estimator.

Let  $\Theta_\delta^*$  denote the compact set

$$\Theta_\delta^* = \left\{ \theta \in \mathbb{R}^{p+q}; \text{ the roots of the polynomials } a_\theta(z) \text{ and } b_\theta(z) \text{ have modulus } \geq 1 + \delta \right\}.$$

We define the set  $\Theta_\delta$  as the cartesian product of  $\Theta_\delta^*$  by  $[d_1, d_2]$ , i.e.  $\Theta_\delta = \Theta_\delta^* \times [d_1, d_2]$ , where  $\delta$  is a strictly positive constant chosen such that  $\theta_0$  belongs to  $\Theta_\delta$ .

The random variable  $\hat{\theta}_n$  is called least squares estimator if it satisfies, almost surely,

$$\hat{\theta}_n = \underset{\theta \in \Theta_\delta}{\operatorname{argmin}} Q_n(\theta), \text{ where } Q_n(\theta) = \frac{1}{n} \sum_{t=1}^n \tilde{\epsilon}_t^2(\theta). \quad (4)$$

Our main results are proven under the following assumptions:

**(A2):** The process  $(\epsilon_t)_{t \in \mathbb{Z}}$  is strictly stationary and ergodic.

The strong consistency of the least squares estimator will be proved under the three above assumptions **((A0), (A1) and (A2))**. For the asymptotic normality of the LSE, additional assumptions are required. It is necessary to assume that  $\theta_0$  is not on the boundary of the parameter space  $\Theta$ .

**(A3):** We have  $\theta_0 \in \overset{\circ}{\Theta}$ , where  $\overset{\circ}{\Theta}$  denotes the interior of  $\Theta$ .

The stationary process  $\epsilon$  is not supposed to be an independent sequence. So one needs to control its dependency by means of its strong mixing coefficients  $\{\alpha_\epsilon(h)\}_{h \in \mathbb{N}}$  defined by

$$\alpha_\epsilon(h) = \sup_{A \in \mathcal{F}_{-\infty}^t, B \in \mathcal{F}_{t+h}^\infty} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

where  $\mathcal{F}_{-\infty}^t = \sigma(\epsilon_u, u \leq t)$  and  $\mathcal{F}_{t+h}^\infty = \sigma(\epsilon_u, u \geq t+h)$ .

We shall need an integrability assumption on the moments of the noise  $\epsilon$  and a summability condition on the strong mixing coefficients  $(\alpha_\epsilon(k))_{k \geq 0}$ .

**(A4):** There exists an integer  $\tau$  such that for some  $\nu \in ]0, 1]$ , we have  $\mathbb{E}|\epsilon_t|^{\tau+\nu} < \infty$  and  $\sum_{h=0}^\infty (h+1)^{k-2} \{\alpha_\epsilon(h)\}^{\frac{\nu}{k+\nu}} < \infty$  for  $k = 1, \dots, \tau$ .

Note that **(A4)** implies the following weak assumption on the joint cumulants of the innovation process  $\epsilon$  (see [DL89], for more details).

**(A4'):** There exists an integer  $\tau \geq 2$  such that  $C_\tau := \sum_{i_1, \dots, i_{\tau-1} \in \mathbb{Z}} |\operatorname{cum}(\epsilon_0, \epsilon_{i_1}, \dots, \epsilon_{i_{\tau-1}})| < \infty$ .

In the above expression,  $\operatorname{cum}(\epsilon_0, \epsilon_{i_1}, \dots, \epsilon_{i_{\tau-1}})$  denotes the  $\tau$ -th order cumulant of the stationary process. Due to the fact that the  $\epsilon_t$ 's are centered, we notice that for fixed  $(i, j, k)$

$$\operatorname{cum}(\epsilon_0, \epsilon_i, \epsilon_j, \epsilon_k) = \mathbb{E}[\epsilon_0 \epsilon_i \epsilon_j \epsilon_k] - \mathbb{E}[\epsilon_0 \epsilon_i] \mathbb{E}[\epsilon_j \epsilon_k] - \mathbb{E}[\epsilon_0 \epsilon_j] \mathbb{E}[\epsilon_i \epsilon_k] - \mathbb{E}[\epsilon_0 \epsilon_k] \mathbb{E}[\epsilon_i \epsilon_j].$$

Assumption **(A4)** is a usual technical hypothesis which is useful when one prove the asymptotic normality (see [FZ98] for example). Let us notice however that we impose a stronger convergence speed for the mixing coefficients than in the works on weak ARMA processes. This is due to the fact that the coefficients in the AR or MA representation of  $\epsilon_t(\theta)$  have no more exponential decay because of the fractional operator (see Subsection 6.1 for details and comments).

As mentioned before, Hypothesis **(A4)** implies **(A4')** which is also a technical assumption usually used in the fractionally integrated ARMA processes framework (see for instance [Sha10c]) or even in an ARMA context (see [FZ07, ZL15]). One remarks that in [Sha10b], the author emphasized that a geometric moment contraction implies **(A4')**. This provides an alternative to strong mixing assumptions but, to our knowledge, there is no relation between this two kinds of hypotheses.

## 2.2. Asymptotic properties

The asymptotic properties of the LSE of the weak FARIMA model are stated in the following two theorems.

**Theorem 1.** (Consistency). Assume that  $(\epsilon_t)_{t \in \mathbb{Z}}$  satisfies (1) and belonging to  $\mathbb{L}^2$ . Let  $(\hat{\theta}_n)_n$  be a sequence of least squares estimators. Under Assumptions (A0), (A1) and (A2), we have

$$\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0.$$

The proof of this theorem is given in Subsection 6.2.

In order to state our asymptotic normality result, we define the function

$$O_n(\theta) = \frac{1}{n} \sum_{t=1}^n \epsilon_t^2(\theta), \quad (5)$$

where the sequence  $(\epsilon_t(\theta))_{t \in \mathbb{Z}}$  is given by (2). We consider the following information matrices

$$I(\theta) = \lim_{n \rightarrow \infty} \text{Var} \left\{ \sqrt{n} \frac{\partial}{\partial \theta} O_n(\theta) \right\} \text{ and } J(\theta) = \lim_{n \rightarrow \infty} \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} O_n(\theta) \right] \text{ a.s.}$$

The existence of these matrices are proved when one demonstrates the following result.

**Theorem 2.** (Asymptotic normality). We assume that  $(\epsilon_t)_{t \in \mathbb{Z}}$  satisfies (1). Under (A0)-(A3) and Assumption (A4) with  $\tau = 4$ . The sequence  $(\sqrt{n}(\hat{\theta}_n - \theta_0))_{n \geq 1}$  has a limiting centered normal distribution with covariance matrix  $\Omega := J^{-1}(\theta_0)I(\theta_0)J^{-1}(\theta_0)$ .

The proof of this theorem is given in Subsection 6.3.

**Remark 1.** Hereafter (see more precisely (52)), we will be able to prove that

$$J(\theta_0) = 2\mathbb{E} \left[ \frac{\partial}{\partial \theta} \epsilon_t(\theta_0) \frac{\partial}{\partial \theta'} \epsilon_t(\theta_0) \right].$$

Thus the matrix  $J(\theta_0)$  has the same expression in the strong and weak FARIMA cases (see Theorem 1 of [Ber95]). On the contrary, the matrix  $I(\theta_0)$  is in general much more complicated in the weak case than in the strong case.

**Remark 2.** In the standard strong FARIMA case, i.e. when (A2) is replaced by the assumption that  $(\epsilon_t)$  is iid, we have  $I(\theta_0) = 2\sigma_\epsilon^2 J(\theta_0)$ . Thus the asymptotic covariance matrix is then reduced as  $\Omega_S := 2\sigma_\epsilon^2 J^{-1}(\theta_0)$ . Generally, when the noise is not an independent sequence, this simplification can not be made and we have  $I(\theta_0) \neq 2\sigma_\epsilon^2 J(\theta_0)$ . The true asymptotic covariance matrix  $\Omega = J^{-1}(\theta_0)I(\theta_0)J^{-1}(\theta_0)$  obtained in the weak FARIMA framework can be very different from  $\Omega_S$ . As a consequence, for the statistical inference on the parameter, the ready-made softwares used to fit FARIMA do not provide a correct estimation of  $\Omega$  for weak FARIMA processes because the standard time series analysis softwares use empirical estimators of  $\Omega_S$ . The problem also holds in the weak ARMA case (see [FZ07] and the references therein). This is why it is interesting to find an estimator of  $\Omega$  which is consistent for both weak and (semi-)strong FARIMA cases.

Based on the above remark, the next subsection deals with two different methods in order find an estimator of  $\Omega$ .



### 3. Estimating the asymptotic variance matrix

For statistical inference problem, the asymptotic variance  $\Omega$  has to be estimated. In particular Theorem 2 can be used to obtain confidence intervals and significance tests for the parameters.

First of all, the matrix  $J(\theta_0)$  can be estimated empirically by the square matrix  $\hat{J}_n$  of order  $p + q + 1$  defined by:

$$\hat{J}_n = \frac{2}{n} \sum_{t=1}^n \left\{ \frac{\partial}{\partial \theta} \tilde{\epsilon}_t(\hat{\theta}_n) \right\} \left\{ \frac{\partial}{\partial \theta'} \tilde{\epsilon}_t(\hat{\theta}_n) \right\}. \quad (6)$$

The convergence of  $\hat{J}_n$  to  $J(\theta_0)$  is classical (see Lemma 8 in Subsection 6.3 for details).

In the standard strong FARIMA case, in view of remark 2, we have  $\hat{\Omega}_S := 2\hat{\sigma}_\epsilon^2 \hat{J}_n^{-1}$  with  $\hat{\sigma}_\epsilon^2 = Q_n(\hat{\theta}_n)$ . Thus  $\hat{\Omega}_S$  is a strongly consistent estimator of  $\Omega_S$ . In the general weak FARIMA case, this estimator is not consistent when  $I(\theta_0) \neq 2\sigma_\epsilon^2 J(\theta_0)$ . So we need a consistent estimator of  $I(\theta_0)$ .

#### 3.1. Estimation of the asymptotic matrix $I(\theta_0)$

For all  $t \in \mathbb{Z}$ , let

$$H_t(\theta_0) = 2\epsilon_t(\theta_0) \frac{\partial}{\partial \theta} \epsilon_t(\theta_0) = \left( 2\epsilon_t(\theta_0) \frac{\partial}{\partial \theta_1} \epsilon_t(\theta_0), \dots, 2\epsilon_t(\theta_0) \frac{\partial}{\partial \theta_{p+q+1}} \epsilon_t(\theta_0) \right)'. \quad (7)$$

We shall see in the proof of Lemma 9 that

$$I(\theta_0) = \lim_{n \rightarrow \infty} \text{Var} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n H_t(\theta_0) \right) = \sum_{h=-\infty}^{\infty} \text{Cov}(H_t(\theta_0), H_{t-h}(\theta_0)).$$

Following the arguments developed in [BMCF12], the matrix  $I(\theta_0)$  can be estimated using Berk's approach (see [Ber74]). More precisely, by interpreting  $I(\theta_0)/2\pi$  as the spectral density of the stationary process  $(H_t(\theta_0))_{t \in \mathbb{Z}}$  evaluated at frequency 0, we can use a parametric autoregressive estimate of the spectral density of  $(H_t(\theta_0))_{t \in \mathbb{Z}}$  in order to estimate the matrix  $I(\theta_0)$ .

For any  $\theta \in \Theta$ ,  $H_t(\theta)$  is a measurable function of  $\{\epsilon_s, s \leq t\}$ . The stationary process  $(H_t(\theta_0))_{t \in \mathbb{Z}}$  admits the following Wold decomposition  $H_t(\theta_0) = u_t + \sum_{k=1}^{\infty} \psi_k u_{t-k}$ , where  $(u_t)_{t \in \mathbb{Z}}$  is a  $(p + q + 1)$ -variate weak white noise with variance matrix  $\Sigma_u$ .

Assume that  $\Sigma_u$  is non-singular, that  $\sum_{k=1}^{\infty} \|\psi_k\| < \infty$ , and that  $\det(I_{p+q+1} + \sum_{k=1}^{\infty} \psi_k z^k) \neq 0$  when  $|z| \leq 1$ . Then  $(H_t(\theta_0))_{t \in \mathbb{Z}}$  admits a weak multivariate AR( $\infty$ ) representation of the form

$$\Phi(L)H_t(\theta_0) := H_t(\theta_0) - \sum_{k=1}^{\infty} \Phi_k H_{t-k}(\theta_0) = u_t, \quad (8)$$

such that  $\sum_{k=1}^{\infty} \|\Phi_k\| < \infty$  and  $\det \{\Phi(z)\} \neq 0$  for all  $|z| \leq 1$ . It is proved in [BM09, Lüt05] that one may find a constant  $K$  and  $0 < \rho < 1$  such that

$$\|\Phi_k\| \leq K \rho^k. \quad (9)$$

Thanks to the previous remarks, the estimation of  $I(\theta_0)$  is therefore based on the following expression

$$I(\theta_0) = \Phi^{-1}(1) \Sigma_u \Phi^{-1}(1).$$



Consider the regression of  $H_t(\theta_0)$  on  $H_{t-1}(\theta_0), \dots, H_{t-r}(\theta_0)$  defined by

$$H_t(\theta_0) = \sum_{k=1}^r \phi_{r,k} H_{t-k}(\theta_0) + u_{r,t}, \quad (10)$$

where  $u_{r,t}$  is uncorrelated with  $H_{t-1}(\theta_0), \dots, H_{t-r}(\theta_0)$ . Since  $H_t(\theta_0)$  is not observable, we introduce  $\hat{H}_t \in \mathbb{R}^{p+q+1}$  obtained by replacing  $\epsilon_t(\theta_0)$  by  $\tilde{\epsilon}_t(\theta_0)$  and  $\theta_0$  by  $\hat{\theta}_n$  in (7):

$$\hat{H}_t = 2\tilde{\epsilon}_t(\hat{\theta}_n) \frac{\partial}{\partial \theta} \hat{\epsilon}_t(\theta_n). \quad (11)$$

Let  $\hat{\phi}_r(z) = I_{p+q+1} - \sum_{k=1}^r \hat{\phi}_{r,k} z^k$ , where  $\hat{\phi}_{r,1}, \dots, \hat{\phi}_{r,r}$  denote the coefficients of the LS regression of  $\hat{H}_t$  on  $\hat{H}_{t-1}, \dots, \hat{H}_{t-r}$ . Let  $\hat{u}_{r,t}$  be the residuals of this regression and let  $\hat{\Sigma}_{\hat{u}_r}$  be the empirical variance (defined in (12) below) of  $\hat{u}_{r,1}, \dots, \hat{u}_{r,r}$ . The LSE of  $\underline{\phi}_r = (\phi_{r,1}, \dots, \phi_{r,r})$  and  $\Sigma_{u_r} = \text{Var}(u_{r,t})$  are given by

$$\underline{\hat{\phi}}_r = \hat{\Sigma}_{\hat{H}, \hat{H}_r}^{-1} \hat{\Sigma}_{\hat{H}_r}^{-1} \text{ and } \hat{\Sigma}_{\hat{u}_r} = \frac{1}{n} \sum_{t=1}^n \left( \hat{H}_t - \underline{\hat{\phi}}_r \hat{H}_{r,t} \right) \left( \hat{H}_t - \underline{\hat{\phi}}_r \hat{H}_{r,t} \right)', \quad (12)$$

where

$$\hat{H}_{r,t} = (\hat{H}'_{t-1}, \dots, \hat{H}'_{t-r})', \quad \hat{\Sigma}_{\hat{H}, \hat{H}_r} = \frac{1}{n} \sum_{t=1}^n \hat{H}_t \hat{H}_{r,t}' \text{ and } \hat{\Sigma}_{\hat{H}_r} = \frac{1}{n} \sum_{t=1}^n \hat{H}_{r,t} \hat{H}_{r,t}',$$

with by convention  $\hat{H}_t = 0$  when  $t \leq 0$ . We assume that  $\hat{\Sigma}_{\hat{H}_r}$  is non-singular (which holds true asymptotically).

In the case of linear processes with independent innovations, Berk (see [Ber74]) has shown that the spectral density can be consistently estimated by fitting autoregressive models of order  $r = r(n)$ , whenever  $r$  tends to infinity and  $r^3/n$  tends to 0 as  $n$  tends to infinity. There are differences with [Ber74]:  $(H_t(\theta_0))_{t \in \mathbb{Z}}$  is multivariate, is not directly observed and is replaced by  $(\hat{H}_t)_{t \in \mathbb{Z}}$ . It is shown that this result remains valid for the multivariate linear process  $(H_t(\theta_0))_{t \in \mathbb{Z}}$  with non-independent innovations (see [BMCF12, BMF11], for references in weak (multivariate) ARMA models). We will extend the results of [BMCF12] to weak FARIMA models.

The asymptotic study of the estimator of  $I(\theta_0)$  using the spectral density method is given in the following theorem.

**Theorem 3.** *We assume (A0)-(A3) and Assumption (A4') with  $\tau = 8$ . In addition, we assume that the process  $(H_t(\theta_0))_{t \in \mathbb{Z}}$  defined in (7) admits a multivariate  $AR(\infty)$  representation (8). Then, the spectral estimator of  $I(\theta_0)$*

$$\hat{I}_n^{\text{SP}} := \hat{\phi}_r^{-1}(1) \hat{\Sigma}_{\hat{u}_r} \hat{\phi}_r^{-1}(1) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} I(\theta_0) = \Phi^{-1}(1) \Sigma_u \Phi^{-1}(1)$$

where  $r$  depends on  $n$  and satisfies  $\lim_{n \rightarrow \infty} r^5(n)/n^{1-2(d_2-d_1)} = 0$  (remind that  $d_0 \in [d_1, d_2] \subset ]0, 1/2[$ ).

The proof of this theorem is given in Subsection 6.4.

A second method to estimate the asymptotic matrix (or rather avoiding estimate it) is proposed in the next subsection.

### 3.2. A self-normalized approach to confidence interval construction in weak FARIMA models

We have seen previously that we may obtain confidence intervals for weak FARIMA model parameters as soon as we can construct a convergent estimator of the variance matrix  $I(\theta_0)$  (see Theorems 2 and 3). The parametric approach based on an autoregressive estimate of the spectral density of  $(H_t(\theta_0))_{t \in \mathbb{Z}}$  that we used before has the drawback of choosing the truncation parameter  $r$  in (10). This choice of the order truncation is often crucial and difficult. So the aim of this section is to avoid such a difficulty.

This section is also of interest because, to our knowledge, it has not been studied for weak FARIMA models. Notable exception is [Sha12] who studied this problem in a short memory case (see Assumption 1 in [Sha12] that implies that the process  $X$  is short-range dependent).

We propose an alternative method to obtain confidence intervals for weak FARIMA models by avoiding the estimation of the asymptotic covariance matrix  $I(\theta_0)$ . It is based on a self-normalization approach used to build a statistic which depends on the true parameter  $\theta_0$  and which is asymptotically distribution-free (see Theorem 1 of [Sha12] for a reference in weak ARMA case). The idea comes from [Lob01] and has been already extended by [BMS18, KL06, Sha10c, Sha10a, Sha12] to more general frameworks. See also [Sha15] for a review on some recent developments on the inference of time series data using the self-normalized approach.

Let us briefly explain the idea of the self-normalization.

By a Taylor expansion of the function  $\partial Q_n(\cdot)/\partial \theta$  around  $\theta_0$ , under (A3), we have

$$0 = \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\hat{\theta}_n) = \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\theta_0) + \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\theta_{n,i,j}^*) \right] \sqrt{n} (\hat{\theta}_n - \theta_0), \quad (13)$$

where the  $\theta_{n,i,j}^*$ 's are between  $\hat{\theta}_n$  and  $\theta_0$ . Using the following equation

$$\sqrt{n} \left( \frac{\partial}{\partial \theta} O_n(\theta_0) - \frac{\partial}{\partial \theta} Q_n(\theta_0) \right) = \sqrt{n} \frac{\partial}{\partial \theta} O_n(\theta_0) + \left\{ \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\theta_{n,i,j}^*) \right] - J(\theta_0) + J(\theta_0) \right\} \sqrt{n} (\hat{\theta}_n - \theta_0),$$

we shall be able to prove that (13) implies that

$$\sqrt{n} \frac{\partial}{\partial \theta} O_n(\theta_0) + J(\theta_0) \sqrt{n} (\hat{\theta}_n - \theta_0) = o_{\mathbb{P}}(1). \quad (14)$$

This is due to the following technical properties:

- the convergence in probability of  $\sqrt{n} \partial Q_n(\theta_0)/\partial \theta$  to  $\sqrt{n} \partial O_n(\theta_0)/\partial \theta$  (see Lemma 5 hereafter),
- the almost-sure convergence of  $[\partial^2 Q_n(\theta_{n,i,j}^*)/\partial \theta_i \partial \theta_j]$  to  $J(\theta_0)$  (see Lemma 8 hereafter),
- the tightness of the sequence  $(\sqrt{n}(\hat{\theta}_n - \theta_0))_n$  (see Theorem 2) and
- the existence and invertibility of the matrix  $J(\theta_0)$  (see Lemma 6 hereafter).

Thus we obtain from (14) that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t + o_{\mathbb{P}}(1),$$

where (remind (7))

$$U_t = -J^{-1}(\theta_0) H_t(\theta_0).$$

At this stage, we do not rely on the classical method that would consist in estimating the asymptotic covariance matrix  $I(\theta_0)$ . We rather try to apply Lemma 1 in [Lob01]. So we need to check that a functional central limit theorem holds for the process  $U := (U_t)_{t \geq 1}$ . For that sake, we define the normalization matrix  $P_{p+q+1,n}$  of  $\mathbb{R}^{(p+q+1) \times (p+q+1)}$  by

$$P_{p+q+1,n} = \frac{1}{n^2} \sum_{t=1}^n \left( \sum_{j=1}^t (U_j - \bar{U}_n) \right) \left( \sum_{j=1}^t (U_j - \bar{U}_n) \right)', \quad (15)$$

where  $\bar{U}_n = (1/n) \sum_{i=1}^n U_i$ . To ensure the invertibility of the normalization matrix  $P_{p+q+1,n}$  (it is the result stated in the next proposition), we need the following technical assumption on the distribution of  $\epsilon_t$ .

**(A5):** The process  $(\epsilon_t)_{t \in \mathbb{Z}}$  has a positive density on some neighborhood of zero.

**Proposition 4.** *Under the assumptions of Theorem 2 and (A5), the matrix  $P_{p+q+1,n}$  is almost surely non singular.*

The proof of this proposition is given in Subsection 6.5.

Let  $(B_m(r))_{r \geq 0}$  be a  $m$ -dimensional Brownian motion starting from 0. For  $m \geq 1$ , we denote by  $\mathcal{U}_m$  the random variable defined by:

$$\mathcal{U}_m = B_m'(1) V_m^{-1} B_m(1), \quad (16)$$

where

$$V_m = \int_0^1 (B_m(r) - r B_m(1)) (B_m(r) - r B_m(1))' dr. \quad (17)$$

The critical values of  $\mathcal{U}_m$  have been tabulated by [Lob01].

The following theorem states the self-normalized asymptotic distribution of the random vector  $\sqrt{n}(\hat{\theta}_n - \theta_0)$ .

**Theorem 5.** *Under the assumptions of Theorem 2 and (A5), we have*

$$n(\hat{\theta}_n - \theta_0)' P_{p+q+1,n}^{-1} (\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{in\ law} \mathcal{U}_{p+q+1}.$$

The proof of this theorem is given in Subsection 6.6.

Of course, the above theorem is useless for practical purpose because the normalization matrix  $P_{p+q+1,n}$  is not observable. This gap will be fixed below when one replaces the matrix  $P_{p+q+1,n}$  by its empirical or observable counterpart

$$\hat{P}_{p+q+1,n} = \frac{1}{n^2} \sum_{t=1}^n \left( \sum_{j=1}^t (\hat{U}_j - \frac{1}{n} \sum_{k=1}^n \hat{U}_k) \right) \left( \sum_{j=1}^t (\hat{U}_j - \frac{1}{n} \sum_{k=1}^n \hat{U}_k) \right)' \quad \text{where } \hat{U}_j = -\hat{J}_n^{-1} \hat{H}_j. \quad (18)$$

The above quantity is observable and we are able to state our Theorem which is the applicable version of Theorem 5.

**Theorem 6.** *Under the assumptions of Theorem 2 and (A5), we have*

$$n(\hat{\theta}_n - \theta_0)' \hat{P}_{p+q+1,n}^{-1} (\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{in\ law} \mathcal{U}_{p+q+1}.$$

The proof of this theorem is given in Subsection 6.7.

At the asymptotic level  $\alpha$ , a joint  $100(1 - \alpha)\%$  confidence region for the elements of  $\theta_0$  is then given by the set of values of the vector  $\theta$  which satisfy the following inequality:

$$n(\hat{\theta}_n - \theta)' \hat{P}_{p+q+1,n}^{-1} (\hat{\theta}_n - \theta) \leq \mathcal{U}_{p+q+1,\alpha},$$

where  $\mathcal{U}_{p+q+1,\alpha}$  is the quantile of order  $1 - \alpha$  for the distribution of  $\mathcal{U}_{p+q+1}$ .

**Corollary 7.** For any  $1 \leq i \leq p + q + 1$ , a  $100(1 - \alpha)\%$  confidence region for  $\theta_0(i)$  is given by the following set:

$$\left\{ x \in \mathbb{R} ; n(\hat{\theta}_n(i) - x)^2 \hat{P}_{p+q+1,n}^{-1}(i, i) \leq \mathcal{U}_{1,\alpha} \right\},$$

where  $\mathcal{U}_{1,\alpha}$  denotes the quantile of order  $1 - \alpha$  of the distribution for  $\mathcal{U}_1$ .

The proof of this corollary is similar to that of Theorem 6 when one restricts ourselves to a one dimensional case.

#### 4. Numerical illustrations

In this section, we investigate the finite sample properties of the asymptotic results that we introduced in this work. For that sake we use Monte Carlo experiments. The numerical illustrations of this section are made with the open source statistical software R (see R Development Core Team, 2017) or (see <http://cran.r-project.org/>).

##### 4.1. Simulation studies and empirical sizes for confidence intervals

We study numerically the behavior of the LSE for FARIMA models of the form

$$(1 - L)^d (X_t - aX_{t-1}) = \epsilon_t - b\epsilon_{t-1}, \quad (19)$$

where the unknown parameter is taken as  $\theta_0 = (a, b, d) = (-0.7, -0.2, 0.4)$ . First we assume that in (19) the innovation process  $(\epsilon_t)_{t \in \mathbb{Z}}$  is an iid centered Gaussian process with common variance 1 which corresponds to the strong FARIMA case. In two other experiments we consider that in (19) the innovation processes  $(\epsilon_t)_{t \in \mathbb{Z}}$  are defined respectively by

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = 0.04 + 0.12\epsilon_{t-1}^2 + 0.85\sigma_{t-1}^2 \end{cases} \quad (20)$$

and

$$\epsilon_t = \eta_t^2 \eta_{t-1}, \quad (21)$$

where  $(\eta_t)_{t \geq 1}$  is a sequence of iid centered Gaussian random variables with variance 1. Note that the innovation process in (21) is not a martingale difference whereas it is the case of the noise defined in (20).

We simulated  $N = 1,000$  independent trajectories of size  $n = 2,000$  of Model (19) in the three following case: the strong Gaussian noise, the semi-strong noise (20) and the weak noise (21).

Figure 1, Figure 2 and Figure 3 compare the distribution of the LSE in these three contexts. The distributions of  $\hat{d}_n$  are similar in the three cases whereas the LSE  $\hat{a}_n$  of  $a$  is more accurate in the weak case than in the strong and semi-strong cases. The distributions of  $\hat{b}_n$  are more accurate

in the strong case than in the weak case. Remark that in the weak case the distributions of  $\hat{b}_n$  are more accurate to the semi-strong ones.

Figure 4 compares standard estimator  $\hat{\Omega}_S = 2\hat{\sigma}_\epsilon^2 \hat{J}_n^{-1}$  and the sandwich estimator  $\hat{\Omega} = \hat{J}_n^{-1} \hat{I}_n^{\text{SP}} \hat{J}_n^{-1}$  of the LSE asymptotic variance  $\Omega$ . We used the spectral estimator  $\hat{I}_n^{\text{SP}}$  defined in Theorem 3. The multivariate AR order  $r$  (see (10)) is automatically selected by AIC (we use the function `VARselect()` of the `vars` R package). In the strong FARIMA case we know that the two estimators are consistent. In view of the two upper subfigures of Figure 4, it seems that the sandwich estimator is less accurate in the strong case. This is not surprising because the sandwich estimator is more robust, in the sense that this estimator remains consistent in the semi-strong and weak FARIMA cases, contrary to the standard estimator (see the middle and bottom subfigures of Figure 4). Figure 5 (resp. Figure 6) presents a zoom of the left(right)-middle and left(right)-bottom panels of Figure 4. It is clear that in the semi-strong or weak case  $n(\hat{a}_n - a)^2$ ,  $n(\hat{b}_n - b)^2$  and  $n(\hat{d}_n - d)^2$  are, respectively, better estimated by  $\hat{J}_n^{-1} \hat{I}_n^{\text{SP}} \hat{J}_n^{-1}(1, 1)$ ,  $\hat{J}_n^{-1} \hat{I}_n^{\text{SP}} \hat{J}_n^{-1}(2, 2)$  and  $\hat{J}_n^{-1} \hat{I}_n^{\text{SP}} \hat{J}_n^{-1}(3, 3)$  (see Figure 6) than by  $2\hat{\sigma}_\epsilon^2 \hat{J}_n^{-1}(1, 1)$ ,  $2\hat{\sigma}_\epsilon^2 \hat{J}_n^{-1}(2, 2)$  and  $2\hat{\sigma}_\epsilon^2 \hat{J}_n^{-1}(3, 3)$  (see Figure 5). The failure of the standard estimator of  $\Omega$  in the weak FARIMA framework may have important consequences in terms of identification or hypothesis testing and validation.

Now we are interested in standard confidence interval and the modified versions proposed in Subsections 3.1 and 3.2. Table 1 displays the empirical sizes in the three previous different FARIMA cases. For the nominal level  $\alpha = 5\%$ , the empirical size over the  $N = 1,000$  independent replications should vary between the significant limits 3.6% and 6.4% with probability 95%. For the nominal level  $\alpha = 1\%$ , the significant limits are 0.3% and 1.7%, and for the nominal level  $\alpha = 10\%$ , they are 8.1% and 11.9%. When the relative rejection frequencies are outside the significant limits, they are displayed in bold type in Table 1. For the strong FARIMA model, all the relative rejection frequencies are inside the significant limits for  $n$  large. For the semi-strong FARIMA model, the relative rejection frequencies of the standard confidence interval are definitely outside the significant limits, contrary to the modified versions proposed. For the weak FARIMA model, only the standard confidence interval of  $\hat{b}_n$  is outside the significant limits when  $n$  increases. As a conclusion, Table 1 confirms the comments done concerning Figure 4.

#### 4.2. Application to real data

We now consider an application to the daily returns of four stock market indices (CAC, DAX, Nikkei and S&P 500). The returns are defined by  $r_t = \log(p_t/p_{t-1})$  where  $p_t$  denotes the price index of the stock market indices at time  $t$ . The observations cover the period from the starting date of each index to February 14, 2019. Figure 7 (resp. Figure 8) plots the closing prices (resp. the returns) of the four stock market indices. Figure 9 shows that the squared returns  $(r_t^2)_{t \geq 1}$  are generally strongly autocorrelated.

In Financial Econometrics the returns are often assumed to be martingale increments. The squares of the returns have often second-order moments close to those of an ARMA(1, 1) which is compatible with a GARCH(1, 1) model for the returns (see [FZ10]). A long-range memory property of the stock market returns series was also largely investigated by [DGE93] (see also [BFGK13], [Pal07] and [BCT96]). The squared returns  $(r_t^2)_{t \geq 1}$  have significant positive autocorrelations at least up to lag 100 (see Figure 9) which confirm the claim that stock market returns have long-term memory (see [DGE93]). In particular the returns  $(r_t)_{t \geq 1}$  process is characterized by substantially more correlation between absolute or squared returns than between the returns themselves.

Now we focus on the dynamics of the squared returns and we fit a FARIMA(1,  $d$ , 1) model to the squares of the 4 daily returns. Denoting by  $(X_t)_{t \geq 1}$  the mean corrected series of the squared

returns, we adjust the following model

$$(1 - L)^d (X_t - aX_{t-1}) = \epsilon_t - b\epsilon_{t-1}.$$

Table 2 displays the LSE of the parameter  $\theta = (a, b, d)$  of each squared of daily returns. The  $p$ -values of the corresponding LSE,  $\hat{\theta}_n = (\hat{a}_n, \hat{b}_n, \hat{d}_n)$  are given in parentheses. The last column presents the estimated residual variance. Note that for all series, the estimated coefficients  $|\hat{a}_n|$  and  $|\hat{b}_n|$  are smaller than one and this is in accordance with our Assumption (A1). We also observe that for all series the estimated long-range dependence coefficients  $\hat{d}_n$  are significant for any reasonable asymptotic level and are inside  $]0, 0.5[$ . We thus think that the assumption (A3) is satisfied and thus our asymptotic normality theorem can be applied. Table 3 then presents for each serie the modified confidence interval at the asymptotic level  $\alpha = 5\%$  for the parameters estimated in Table 2.

## 5. Conclusion

Taking into account the possible lack of independence of the error terms, we show in this paper that we can fit FARIMA representations of a wide class of nonlinear long memory times series. This is possible thanks to our theoretical results and it is illustrated in our real cases and simulations studies.

This standard methodology (when the noise is supposed to be iid), in particular the significance tests on the parameters, needs however to be adapted to take into account the possible lack of independence of the errors terms. A first step has been done thanks to our results on the confidence intervals. In future works, we intent to study how the existing identification (see [BM12], [BMK16]) and diagnostic checking (see [BMS18], [FRZ05]) procedures should be adapted in the presence of long-range dependence framework and dependent noise.

## 6. Proofs

In all our proofs,  $K$  is a strictly positive constant that may vary from line to line.

### 6.1. Preliminary results

In this subsection, we shall give some results on estimations of the coefficient of formal power series that will arise in our study. Some of them are well know on some others are new to our knowledge. We will make some precise comments hereafter.

We begin by recalling the following properties on power series. If for  $|z| \leq R$ , the power series  $f(z) = \sum_{i \geq 0} a_i z^i$  and  $g(z) = \sum_{i \geq 0} b_i z^i$  are well defined, then one has  $(fg)(z) = \sum_{i \geq 0} c_i z^i$  is also well defined for  $|z| \leq R$  with the sequence  $(c_i)_{i \geq 0}$  which is given by  $c = a * b$  where  $*$  denotes the convolution product between  $a$  and  $b$  defined by  $c_i = \sum_{k=0}^i a_k b_{i-k} = \sum_{k=0}^i a_{i-k} b_k$ . We will make use of the Young inequality that states that if the sequence  $a \in \ell^p$  and  $b \in \ell^q$  and such that  $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$  with  $1 \leq p, q, r \leq \infty$ , then

$$\|a * b\|_{\ell^r} \leq \|a\|_{\ell^p} \times \|b\|_{\ell^q}.$$

Now we come back to the power series that arise in our context. Remind that for the true value of the parameter,

$$a_{\theta_0}(L)(1 - L)^{d_0} X_t = b_{\theta_0}(L) \epsilon_t. \quad (22)$$

Thanks to the assumptions on the moving average polynomials  $b_\theta$  and the autoregressive polynomials  $a_\theta$ , the power series  $a_\theta^{-1}$  and  $b_\theta^{-1}$  are well defined.

Thus the functions  $\epsilon_t(\theta)$  defined in (2) can be written as

$$\epsilon_t(\theta) = b_\theta^{-1}(L)a_\theta(L)(1-L)^d X_t \quad (23)$$

$$= b_\theta^{-1}(L)a_\theta(L)(1-L)^{d-d_0} a_{\theta_0}^{-1}(L)b_{\theta_0}(L)\epsilon_t \quad (24)$$

and if we denote  $\gamma(\theta) = (\gamma_i(\theta))_{i \geq 0}$  the sequence of coefficients of the power series  $b_\theta^{-1}(z)a_\theta(z)(1-z)^d$  (which is absolutely convergent for at least for  $|z| \leq 1$ ), we may write for all  $t \in \mathbb{Z}$ :

$$\epsilon_t(\theta) = \sum_{i \geq 0} \gamma_i(\theta) X_{t-i}, \quad (25)$$

In the same way, by (23) one has

$$X_t = (1-L)^{-d} a_\theta^{-1}(L)b_\theta(L)\epsilon_t(\theta)$$

and if we denote  $\eta(\theta) = (\eta_i(\theta))_{i \geq 0}$  the coefficients of the power series  $(1-z)^{-d} a_\theta^{-1}(z)b_\theta(z)$  one has

$$X_t = \sum_{i \geq 0} \eta_i(\theta) \epsilon_{t-i}(\theta). \quad (26)$$

We strength the fact that  $\gamma_0(\theta) = \eta_0(\theta) = 1$  for all  $\theta$ .

For large  $j$ , [HTSC99] have shown that uniformly in  $\theta$  the sequences  $\gamma(\theta)$  and  $\eta(\theta)$  satisfy

$$\frac{\partial^k \gamma_j(\theta)}{\partial \theta_{i_1} \cdots \partial \theta_{i_k}} = O\left(j^{-1-d} \{\log(j)\}^k\right), \text{ for } k = 0, 1, 2, 3, \quad (27)$$

and

$$\frac{\partial^k \eta_j(\theta)}{\partial \theta_{i_1} \cdots \partial \theta_{i_k}} = O\left(j^{-1+d} \{\log(j)\}^k\right), \text{ for } k = 0, 1, 2, 3. \quad (28)$$

Note that, in view of (25), (26) and (27), for all  $\theta \in \Theta_\delta$ ,  $\epsilon_t(\theta)$  belongs to  $\mathbb{L}^2$ , that  $(\epsilon_t(\theta))_{t \in \mathbb{Z}}$  is an ergodic sequence and that, for all  $t \in \mathbb{Z}$ , the function  $\epsilon_t(\cdot)$  is a continuous function.

One difficulty that has to be addressed is that (25) includes the infinite past  $(X_{t-i})_{i \geq 0}$  whereas only a finite number of observations  $(X_t)_{1 \leq t \leq n}$  are available to compute the estimators defined in (4). The simplest solution is truncation which amounts to setting all unobserved values equal to zero. Thus, for all  $\theta \in \Theta$  and  $1 \leq t \leq n$  one defines

$$\tilde{\epsilon}_t(\theta) = \sum_{i=0}^{t-1} \gamma_i(\theta) X_{t-i} = \sum_{i \geq 0} \gamma_i^t(\theta) X_{t-i} \quad (29)$$

where the truncated sequence  $\gamma^t(\theta) = (\gamma_i^t(\theta))_{i \geq 0}$  is defined by

$$\gamma_i^t(\theta) = \begin{cases} \gamma_i(\theta) & \text{if } 0 \leq i \leq t-1, \\ 0 & \text{otherwise.} \end{cases}$$

Since our assumptions are made on the noise in (1), it will be useful to express the random variables  $\epsilon_t(\theta)$  and its partial derivatives with respect to  $\theta$ , as a function of  $(\epsilon_{t-i})_{i \geq 0}$ .



From (24), there exists a sequence  $\lambda(\theta) = (\lambda_i(\theta))_{i \geq 0}$  such that

$$\epsilon_t(\theta) = \sum_{i=0}^{\infty} \lambda_i(\theta) \epsilon_{t-i} \quad (30)$$

where the sequence  $\lambda(\theta)$  is given by the sequence of the coefficients of the power series  $b_\theta^{-1}(z)a_\theta(z)(1-z)^{d-d_0}a_{\theta_0}^{-1}(z)b_{\theta_0}(z)$ . Consequently  $\lambda(\theta) = \gamma(\theta) * \eta(\theta_0)$  or, equivalently,

$$\lambda_i(\theta) = \sum_{j=0}^i \gamma_j(\theta) \eta_{i-j}(\theta_0). \quad (31)$$

We proceed in the same way as regard to the derivatives of  $\epsilon_t(\theta)$ . More precisely, for any  $\theta \in \Theta$ ,  $t \in \mathbb{Z}$  and  $1 \leq k, l \leq p + q + 1$  there exists sequences  $\dot{\lambda}_k(\theta) = (\dot{\lambda}_{i,k}(\theta))_{i \geq 1}$  and  $\ddot{\lambda}_{k,l}(\theta) = (\ddot{\lambda}_{i,k,l}(\theta))_{i \geq 1}$  such that

$$\frac{\partial \epsilon_t(\theta)}{\partial \theta_k} = \sum_{i=1}^{\infty} \dot{\lambda}_{i,k}(\theta) \epsilon_{t-i} \quad (32)$$

$$\frac{\partial^2 \epsilon_t(\theta)}{\partial \theta_k \partial \theta_l} = \sum_{i=1}^{\infty} \ddot{\lambda}_{i,k,l}(\theta) \epsilon_{t-i}. \quad (33)$$

Of course it holds that  $\dot{\lambda}_k(\theta) = \frac{\partial \gamma(\theta)}{\partial \theta_k} * \eta(\theta_0)$  and  $\ddot{\lambda}_{k,l}(\theta) = \frac{\partial^2 \gamma(\theta)}{\partial \theta_k \partial \theta_l} * \eta(\theta_0)$ .

Similarly we have

$$\tilde{\epsilon}_t(\theta) = \sum_{i=0}^{\infty} \lambda_i^t(\theta) \epsilon_{t-i}, \quad (34)$$

$$\frac{\partial \tilde{\epsilon}_t(\theta)}{\partial \theta_k} = \sum_{i=1}^{\infty} \dot{\lambda}_{i,k}^t(\theta) \epsilon_{t-i} \quad (35)$$

$$\frac{\partial^2 \tilde{\epsilon}_t(\theta)}{\partial \theta_k \partial \theta_l} = \sum_{i=1}^{\infty} \ddot{\lambda}_{i,k,l}^t(\theta) \epsilon_{t-i}, \quad (36)$$

where  $\lambda^t(\theta) = \gamma^t(\theta) * \eta(\theta_0)$ ,  $\dot{\lambda}_k^t(\theta) = \frac{\partial \gamma^t(\theta)}{\partial \theta_k} * \eta(\theta_0)$  and  $\ddot{\lambda}_{k,l}^t(\theta) = \frac{\partial^2 \gamma^t(\theta)}{\partial \theta_k \partial \theta_l} * \eta(\theta_0)$ .

In order to handle the truncation error  $\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)$ , one needs informations on the sequence  $\lambda(\theta) - \lambda^t(\theta)$ . This is the purpose on the following lemma.

**Lemma 1.** For  $2 \leq r \leq \infty$ ,  $1 \leq k \leq p + q + 1$  and  $\theta \in \Theta$ , we have

$$\| \lambda(\theta) - \lambda^t(\theta) \|_{\ell^r} = O \left( t^{-1+\frac{1}{r}-(d-d_0)} \right)$$

and

$$\| \dot{\lambda}_k(\theta) - \dot{\lambda}_k^t(\theta) \|_{\ell^r} = O \left( t^{-1+\frac{1}{r}-(d-d_0)} \right).$$

*Proof.* We have

$$\lambda(\theta) - \lambda^t(\theta) = (\gamma(\theta) - \gamma^t(\theta)) * \eta(\theta_0).$$

In view of (28), the sequence  $\eta(\theta_0)$  belongs to  $\ell^q$  for any  $q > 1/(1 - d_0)$ . Young's inequality for convolution yields that for all  $r \geq 2$

$$\|\lambda(\theta) - \lambda^t(\theta)\|_{\ell^r} \leq \|\gamma(\theta) - \gamma^t(\theta)\|_{\ell^p} \|\eta(\theta_0)\|_{\ell^q} \quad (37)$$

with  $q = (1 - (d_0 + \beta))^{-1} > 1/(1 - d_0)$  and  $p = r/(1 + r(d_0 + \beta))$ , for some  $\beta > 0$  sufficiently small. Thus there exists  $K$  such that  $\|\eta(\theta_0)\|_{\ell^q} \leq K$ . Since for any  $j \geq 0$ ,

$$\gamma_j(\theta) - \gamma_j^t(\theta) = \begin{cases} 0 & \text{if } 0 \leq j \leq t-1 \\ \gamma_j(\theta) & \text{otherwise} \end{cases}$$

we obtain using (27) that

$$\begin{aligned} \|\lambda(\theta) - \lambda^t(\theta)\|_{\ell^r} &\leq K \left( \sum_{k=0}^{\infty} |\gamma_k(\theta) - \gamma_k^t(\theta)|^p \right)^{1/p} \\ &\leq K \left( \sum_{k=t}^{\infty} |\gamma_k(\theta)|^p \right)^{1/p} \\ &\leq K \left( \sum_{k=t}^{\infty} \frac{1}{k^{p+pd}} \right)^{1/p} \\ &\leq K \left( \int_t^{\infty} \frac{1}{x^{p+pd}} dx \right)^{1/p} \\ &\leq K t^{-1-d+\frac{1}{p}} \\ &\leq K t^{-1+\frac{1}{r}-(d-d_0)+\beta}, \end{aligned}$$

where the constant  $K$  varies from line to line. The conclusion follows by tending  $\beta$  to 0.

The second point of the lemma is shown in the same way as the first. This is because from (27), the coefficient  $\partial \gamma_j(\theta)/\partial \theta_k = O(j^{-1-d+\zeta})$  for any small enough  $\zeta > 0$ . The proof of the lemma is then complete.  $\square$

**Remark 3.** Taking  $r = \infty$  in the above lemma implies that the sequence  $\dot{\lambda}_k(\theta_0) - \dot{\lambda}_k^t(\theta_0)$  is bounded and more precicely there exists  $K$  such that

$$\sup_{j \geq 1} \left| \dot{\lambda}_{j,k}(\theta_0) - \dot{\lambda}_{j,k}^t(\theta_0) \right| \leq \frac{K}{t} \quad (38)$$

for any  $t$  and any  $1 \leq k \leq p + q + 1$ .

One shall also need the following lemmas.

**Lemma 2.** For any  $2 \leq r \leq \infty$ ,  $1 \leq k \leq p + q + 1$  and  $\theta \in \Theta$ , there exists a constant  $K$  such that we have

$$\|\dot{\lambda}_k^t(\theta)\|_{\ell^r} \leq K.$$

*Proof.* The proof follows the same arguments than the proof of Lemma 1.  $\square$

**Lemma 3.** There exists a constant  $K$  such that we have

$$\left| \dot{\lambda}_{i,k}(\theta_0) \right| \leq \frac{K}{i}. \quad (39)$$

*Proof.* For  $1 \leq k \leq p + q + 1$ , the sequence  $\dot{\lambda}_k(\theta) = (\dot{\lambda}_{i,k}(\theta))_{i \geq 1}$  is in fact the sequence of the coefficients in the power series of

$$\frac{\partial}{\partial \theta_k} \left( b_{\theta}^{-1}(z) a_{\theta}(z) (1-z)^{d-d_0} a_{\theta_0}^{-1}(z) b_{\theta_0}(z) \right).$$

Thus  $\dot{\lambda}_{i,k}(\theta_0)$  is the  $i$ -th coefficient taken in  $\theta = \theta_0$ . There are three cases.

◇  $k = 1, \dots, p$ :

Since

$$\frac{\partial}{\partial \theta_k} \left( b_{\theta}^{-1}(z) a_{\theta}(z) (1-z)^{d-d_0} a_{\theta_0}^{-1}(z) b_{\theta_0}(z) \right) = -b_{\theta}^{-1}(z) z^k (1-z)^{d-d_0} a_{\theta_0}^{-1}(z) b_{\theta_0}(z),$$

we deduce that  $\dot{\lambda}_{i,k}(\theta_0)$  is the  $i$ -th coefficient of  $z^k a_{\theta_0}^{-1}(z)$  which satisfies  $\dot{\lambda}_{i,k}(\theta_0) \leq K \rho^i$  for some  $0 < \rho < 1$  (see [FZ98] for example).

◇  $k = p + 1, \dots, p + q$ :

We have

$$\frac{\partial}{\partial \theta_k} \left( b_{\theta}^{-1}(z) a_{\theta}(z) (1-z)^{d-d_0} a_{\theta_0}^{-1}(z) b_{\theta_0}(z) \right) = \left( \frac{\partial}{\partial \theta_k} b_{\theta}^{-1}(z) \right) a_{\theta}(z) (1-z)^{d-d_0} a_{\theta_0}^{-1}(z) b_{\theta_0}(z)$$

and consequently  $\dot{\lambda}_{i,k}(\theta_0)$  is the  $i$ -th coefficient of  $(\frac{\partial}{\partial \theta_k} b_{\theta}^{-1}(z)) b_{\theta_0}(z)$  which also satisfies  $\dot{\lambda}_{i,k}(\theta_0) \leq K \rho^i$  (see [FZ98]).

The last case will not be a consequence of the usual works on ARMA processes.

◇  $k = p + q + 1$ :

In this case,  $\theta_k = d$  and so we have

$$\frac{\partial}{\partial \theta_k} \left( b_{\theta}^{-1}(z) a_{\theta}(z) (1-z)^{d-d_0} a_{\theta_0}^{-1}(z) b_{\theta_0}(z) \right) = b_{\theta}^{-1}(z) a_{\theta}(z) \ln(1-z) (1-z)^{d-d_0} a_{\theta_0}^{-1}(z) b_{\theta_0}(z)$$

and consequently  $\dot{\lambda}_{i,k}(\theta_0)$  is the  $i$ -th coefficient of  $\ln(1-z)$  which is equal to  $-1/i$ .

The three above cases imply the expected result.  $\square$

## 6.2. Proof of Theorem 1

We can follow from line to line the proof of Theorem 1 in [FZ98]. The only difference relies on the following Lemma in which it is stated that the choice of the initial values has no influence on the estimation. Its proof is completely different from the one done in [FZ98] because we do not have the same speed of convergence.

**Lemma 4.** *Under the assumptions of Theorem 1, we have almost surely*

$$\lim_{t \rightarrow \infty} \sup_{\theta \in \Theta_{\delta}} |\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)| = 0. \quad (40)$$

*Proof.* From (25) and (29), for all  $\theta \in \Theta_{\delta}$  and all  $t \in \mathbb{Z}$ , we have

$$\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta) = \sum_{j \geq 0} \gamma_j(\theta) X_{t-j} - \sum_{j=0}^{t-1} \gamma_j(\theta) X_{t-j} = \sum_{j \geq t} \gamma_j(\theta) X_{t-j} = \sum_{k \geq 0} \gamma_{t+k}(\theta) X_{-k}.$$

Recall that for any sequence  $(Y_n)_n$  of random variables it holds that

$$Y_n \xrightarrow[n \rightarrow \infty]{\text{a.s.}} Y \Leftrightarrow \sup_{k \geq n} |Y_k - Y| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

Hence  $\sup_{\theta \in \Theta_\delta} |\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)|$  converges almost surely to 0 as soon as  $\sup_{k \geq t} \sup_{\theta \in \Theta_\delta} |\epsilon_k(\theta) - \tilde{\epsilon}_k(\theta)|$  converges in probability to 0. In view of (27), for all  $\beta > 0$  and for large  $t$  we have

$$\begin{aligned} \mathbb{P} \left( \sup_{k \geq t} \sup_{\theta \in \Theta_\delta} |\epsilon_k(\theta) - \tilde{\epsilon}_k(\theta)| > \beta \right) &= \mathbb{P} \left( \sup_{k \geq t} \sup_{\theta \in \Theta_\delta} \left| \sum_{j \geq 0} \gamma_{k+j}(\theta) X_{-j} \right| > \beta \right) \\ &\leq \mathbb{P} \left( \sum_{j \geq 0} \sup_{k \geq t} \sup_{\theta \in \Theta_\delta} |\gamma_{k+j}(\theta)| |X_{-j}| > \beta \right) \\ &\leq \frac{K}{\beta} \left( \sup_{t \in \mathbb{Z}} \mathbb{E} |X_t| \right) \sum_{j \geq 0} \left( \frac{1}{t+j} \right)^{1+d_1}, \\ &\leq \frac{K}{\beta d_1} (t-1)^{-d_1} \xrightarrow[t \rightarrow \infty]{} 0 \end{aligned}$$

and (40) is proved. □

### 6.3. Proof of Theorem 2

By a Taylor expansion of the function  $\partial Q_n(\cdot)/\partial \theta$  around  $\theta_0$  and under (A3), we have

$$0 = \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\hat{\theta}_n) = \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\theta_0) + \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\theta_{n,i,j}^*) \right] \sqrt{n} (\hat{\theta}_n - \theta_0), \quad (41)$$

where the  $\theta_{n,i,j}^*$ 's are between  $\hat{\theta}_n$  and  $\theta_0$ . The equation (41) can be rewritten in the form:

$$\sqrt{n} \frac{\partial}{\partial \theta} O_n(\theta_0) - \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\theta_0) = \sqrt{n} \frac{\partial}{\partial \theta} O_n(\theta_0) + \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\theta_{n,i,j}^*) \right] \sqrt{n} (\hat{\theta}_n - \theta_0). \quad (42)$$

Under the assumptions of Theorem 2, it will be shown respectively in Lemma 5 and Lemma 8 that

$$\sqrt{n} \frac{\partial}{\partial \theta} O_n(\theta_0) - \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\theta_0) = o_{\mathbb{P}}(1),$$

and

$$\lim_{n \rightarrow \infty} \left\{ \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\theta_{n,i,j}^*) \right] - J(\theta_0) \right\} = 0 \quad \text{a.s.}$$

As a consequence, the asymptotic normality of  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  will be a consequence of the one of  $\sqrt{n} \partial / \partial \theta O_n(\theta_0)$ .

**Lemma 5.** For  $1 \leq k \leq p+q+1$ , under the assumptions of Theorem 2, we have

$$\sqrt{n} \left( \frac{\partial}{\partial \theta_k} Q_n(\theta_0) - \frac{\partial}{\partial \theta_k} O_n(\theta_0) \right) = o_{\mathbb{P}}(1). \quad (43)$$

*Proof.* Throughout this proof,  $\theta = (\theta_1, \dots, \theta_{p+q}, d)' \in \Theta_\delta$  is such that  $d_0 < d \leq d_2$  where  $d_2$  is the upper bound of the support of the long-range dependence parameter  $d_0$ .

The proof is quite long so we divide it in several steps.

◇ **Step 1: preliminaries** For  $1 \leq k \leq p + q + 1$  we have

$$\begin{aligned}
 \sqrt{n} \frac{\partial}{\partial \theta_k} Q_n(\theta_0) &= \frac{2}{\sqrt{n}} \sum_{t=1}^n \tilde{\epsilon}_t(\theta_0) \frac{\partial}{\partial \theta_k} \tilde{\epsilon}_t(\theta_0) \\
 &= \frac{2}{\sqrt{n}} \sum_{t=1}^n (\tilde{\epsilon}_t(\theta_0) - \tilde{\epsilon}_t(\theta)) \frac{\partial}{\partial \theta_k} \tilde{\epsilon}_t(\theta_0) + \frac{2}{\sqrt{n}} \sum_{t=1}^n (\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta)) \frac{\partial}{\partial \theta_k} \tilde{\epsilon}_t(\theta_0) \\
 &\quad + \frac{2}{\sqrt{n}} \sum_{t=1}^n (\epsilon_t(\theta) - \epsilon_t(\theta_0)) \frac{\partial}{\partial \theta_k} \tilde{\epsilon}_t(\theta_0) + \frac{2}{\sqrt{n}} \sum_{t=1}^n \epsilon_t(\theta_0) \left( \frac{\partial}{\partial \theta_k} \tilde{\epsilon}_t(\theta_0) - \frac{\partial}{\partial \theta_k} \epsilon_t(\theta_0) \right) \\
 &\quad + \frac{2}{\sqrt{n}} \sum_{t=1}^n \epsilon_t(\theta_0) \frac{\partial}{\partial \theta_k} \epsilon_t(\theta_0) \\
 &= \Delta_{n,1}^k(\theta) + \Delta_{n,2}^k(\theta) + \Delta_{n,3}^k(\theta) + \Delta_{n,4}^k(\theta_0) + \sqrt{n} \frac{\partial}{\partial \theta_k} O_n(\theta_0), \tag{44}
 \end{aligned}$$

where

$$\begin{aligned}
 \Delta_{n,1}^k(\theta) &= \frac{2}{\sqrt{n}} \sum_{t=1}^n (\tilde{\epsilon}_t(\theta_0) - \tilde{\epsilon}_t(\theta)) \frac{\partial}{\partial \theta_k} \tilde{\epsilon}_t(\theta_0), \\
 \Delta_{n,2}^k(\theta) &= \frac{2}{\sqrt{n}} \sum_{t=1}^n (\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta)) \frac{\partial}{\partial \theta_k} \tilde{\epsilon}_t(\theta_0), \\
 \Delta_{n,3}^k(\theta) &= \frac{2}{\sqrt{n}} \sum_{t=1}^n (\epsilon_t(\theta) - \epsilon_t(\theta_0)) \frac{\partial}{\partial \theta_k} \tilde{\epsilon}_t(\theta_0) \\
 \Delta_{n,4}^k(\theta_0) &= \frac{2}{\sqrt{n}} \sum_{t=1}^n \epsilon_t(\theta_0) \left( \frac{\partial}{\partial \theta_k} \tilde{\epsilon}_t(\theta_0) - \frac{\partial}{\partial \theta_k} \epsilon_t(\theta_0) \right).
 \end{aligned}$$

Using (30) and (34), the fourth term  $\Delta_{n,4}^k(\theta_0)$  can be rewritten in the form:

$$\Delta_{n,4}^k(\theta_0) = \frac{2}{\sqrt{n}} \sum_{t=1}^n \sum_{j=1}^{\infty} \left\{ \dot{\lambda}_{j,k}^t(\theta_0) - \dot{\lambda}_{j,k}(\theta_0) \right\} \epsilon_t \epsilon_{t-j}. \tag{45}$$

Therefore, if we prove that the three sequences of random variables  $(\Delta_{n,1}^k(\theta) + \Delta_{n,3}^k(\theta))_n$ ,  $(\Delta_{n,2}^k(\theta))_n$  and  $(\Delta_{n,4}^k(\theta_0))_n$  converge in probability towards 0, then (43) will be true.

◇ **Step 2: convergence in probability of  $(\Delta_{n,4}^k(\theta_0))_n$  to 0**

For simplicity, in this step we denote in the sequel by  $\dot{\lambda}_{j,k}$  the coefficient  $\dot{\lambda}_{j,k}(\theta_0)$  and by  $\dot{\lambda}_{j,k}^t$  the coefficient  $\dot{\lambda}_{j,k}^t(\theta_0)$ . Let  $\varrho(\cdot, \cdot)$  be the function defined for  $1 \leq t, s \leq n$  by

$$\varrho(t, s) = \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \left\{ \dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^t \right\} \left\{ \dot{\lambda}_{j_2,k} - \dot{\lambda}_{j_2,k}^s \right\} \mathbb{E} [\epsilon_t \epsilon_{t-j_1} \epsilon_s \epsilon_{s-j_2}].$$

For all  $\beta > 0$ , using the symmetry of the function  $\varrho(t, s)$ , we obtain that

$$\begin{aligned} \mathbb{P}\left(\left|\Delta_{n,4}^k(\theta_0)\right| \geq \beta\right) &\leq \frac{4}{n\beta^2} \mathbb{E}\left[\left(\sum_{t=1}^n \sum_{j=1}^{\infty} \left\{\dot{\lambda}_{j,k} - \dot{\lambda}_{j,k}^t\right\} \epsilon_t \epsilon_{t-j}\right)^2\right] \\ &\leq \frac{4}{n\beta^2} \sum_{t=1}^n \sum_{s=1}^n \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \left\{\dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^t\right\} \left\{\dot{\lambda}_{j_2,k} - \dot{\lambda}_{j_2,k}^s\right\} \mathbb{E}[\epsilon_t \epsilon_{t-j_1} \epsilon_s \epsilon_{s-j_2}] \\ &\leq \frac{8}{n\beta^2} \sum_{t=1}^n \sum_{s=1}^t \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \left\{\dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^t\right\} \left\{\dot{\lambda}_{j_2,k} - \dot{\lambda}_{j_2,k}^s\right\} \mathbb{E}[\epsilon_t \epsilon_{t-j_1} \epsilon_s \epsilon_{s-j_2}]. \end{aligned}$$

By the stationarity of  $(\epsilon_t)_{t \in \mathbb{Z}}$  which is assumed in **(A2)**, we have

$$\begin{aligned} \mathbb{E}[\epsilon_t \epsilon_{t-j_1} \epsilon_s \epsilon_{s-j_2}] &= \text{cum}(\epsilon_0, \epsilon_{-j_1}, \epsilon_{s-t}, \epsilon_{s-t-j_2}) + \mathbb{E}[\epsilon_0 \epsilon_{-j_1}] \mathbb{E}[\epsilon_{s-t} \epsilon_{s-t-j_2}] + \mathbb{E}[\epsilon_0 \epsilon_{s-t}] \mathbb{E}[\epsilon_{-j_1} \epsilon_{s-t-j_2}] \\ &\quad + \mathbb{E}[\epsilon_0 \epsilon_{s-t-j_2}] \mathbb{E}[\epsilon_{-j_1} \epsilon_{s-t}]. \end{aligned}$$

Since the noise is not correlated, we deduce that  $\mathbb{E}[\epsilon_0 \epsilon_{-j_1}] = 0$  and  $\mathbb{E}[\epsilon_0 \epsilon_{s-t-j_2}] = 0$  for  $1 \leq j_1, j_2$  and  $s \leq t$ . Consequently we obtain

$$\begin{aligned} \mathbb{P}\left(\left|\Delta_{n,4}^k(\theta_0)\right| \geq \beta\right) &\leq \frac{8}{n\beta^2} \sum_{t=1}^n \sum_{s=1}^t \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \sup_{j_1 \geq 1} \left|\dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^t\right| \left|\dot{\lambda}_{j_2,k} - \dot{\lambda}_{j_2,k}^s\right| |\text{cum}(\epsilon_0, \epsilon_{-j_1}, \epsilon_{s-t}, \epsilon_{s-t-j_2})| \\ &\quad + \frac{8}{n\beta^2} \sum_{t=1}^n \sum_{s=1}^t \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \left|\dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^t\right| \left|\dot{\lambda}_{j_2,k} - \dot{\lambda}_{j_2,k}^s\right| |\mathbb{E}[\epsilon_0 \epsilon_{s-t}] \mathbb{E}[\epsilon_{-j_1} \epsilon_{s-t-j_2}]| \end{aligned} \quad (46)$$

If

$$\sum_{s=1}^t \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \sup_{j_1 \geq 1} \left|\dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^t\right| \left|\dot{\lambda}_{j_2,k} - \dot{\lambda}_{j_2,k}^s\right| |\text{cum}(\epsilon_0, \epsilon_{-j_1}, \epsilon_{s-t}, \epsilon_{s-t-j_2})| \xrightarrow{t \rightarrow \infty} 0, \quad (47)$$

Cesàro's Lemma implies that the first term in the right hand side of (46) tends to 0. Thanks to Lemma 1 applied with  $r = \infty$  (or see Remark 3) and Assumption **(A4')** with  $\tau = 4$ , we obtain that

$$\begin{aligned} &\sum_{s=1}^t \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \sup_{j_1 \geq 1} \left|\dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^t\right| \left|\dot{\lambda}_{j_2,k} - \dot{\lambda}_{j_2,k}^s\right| |\text{cum}(\epsilon_0, \epsilon_{-j_1}, \epsilon_{s-t}, \epsilon_{s-t-j_2})| \\ &\leq \frac{K}{t} \sum_{s=1}^t \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} |\text{cum}(\epsilon_0, \epsilon_{-j_1}, \epsilon_{s-t}, \epsilon_{s-t-j_2})| \\ &\leq \frac{K}{t} \sum_{s=-\infty}^{\infty} \sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} |\text{cum}(\epsilon_0, \epsilon_s, \epsilon_{j_1}, \epsilon_{j_2})| \xrightarrow{t \rightarrow \infty} 0, \end{aligned}$$

hence (47) holds true. Concerning the second term of right hand side of the inequality (46), we have

$$\begin{aligned}
& \frac{8}{n\beta^2} \sum_{t=1}^n \sum_{s=1}^t \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \left| \dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^t \right| \left| \dot{\lambda}_{j_2,k} - \dot{\lambda}_{j_2,k}^s \right| |\mathbb{E}[\epsilon_0 \epsilon_{s-t}] \mathbb{E}[\epsilon_{-j_1} \epsilon_{s-t-j_2}]| \\
&= \frac{8\sigma_\epsilon^2}{n\beta^2} \sum_{t=1}^n \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \left| \dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^t \right| \left| \dot{\lambda}_{j_2,k} - \dot{\lambda}_{j_2,k}^t \right| |\mathbb{E}[\epsilon_{-j_1} \epsilon_{-j_2}]| \\
&= \frac{8\sigma_\epsilon^4}{n\beta^2} \sum_{t=1}^n \sum_{j_1=1}^{\infty} \left| \dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^t \right|^2 \\
&= \frac{8\sigma_\epsilon^4}{n\beta^2} \sum_{t=1}^n \left\| \dot{\lambda}_k - \dot{\lambda}_k^t \right\|_{\ell^2}^2 \\
&\leq \frac{K}{\beta^2} \frac{1}{n} \sum_{t=1}^n \frac{1}{t} \xrightarrow{n \rightarrow \infty} 0
\end{aligned}$$

where we have used the fact that the noise is not correlated, Lemma 1 with  $r = 2$  and Cesàro's Lemma. This ends Step 2.

◇ **Step 3:**  $(\Delta_{n,2}^k(\theta))_n$  converges in probability to 0

For all  $\beta > 0$ , we have

$$\mathbb{P} \left( \left| \Delta_{n,2}^k(\theta) \right| \geq \beta \right) \leq \frac{2}{\beta\sqrt{n}} \sum_{t=1}^n \left\| \tilde{\epsilon}_t(\theta) - \epsilon_t(\theta) \right\|_{\mathbb{L}^2} \left\| \frac{\partial}{\partial \theta_k} \tilde{\epsilon}_t(\theta_0) \right\|_{\mathbb{L}^2}.$$

First, using Lemma 2, we have

$$\begin{aligned}
\left\| \frac{\partial}{\partial \theta_k} \tilde{\epsilon}_t(\theta_0) \right\|_{\mathbb{L}^2}^2 &= \mathbb{E} \left[ \left( \sum_{i=1}^{\infty} \dot{\lambda}_{i,k}^t(\theta_0) \epsilon_{t-i} \right)^2 \right] \\
&= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \dot{\lambda}_{i,k}^t(\theta_0) \dot{\lambda}_{j,k}^t(\theta_0) \mathbb{E}[\epsilon_{t-i} \epsilon_{t-j}] \\
&= \sigma_\epsilon^2 \sum_{i=1}^{\infty} \left\{ \dot{\lambda}_{i,k}^t(\theta_0) \right\}^2 \\
&\leq K.
\end{aligned} \tag{48}$$



In view of (30), (34) and (48), we may write

$$\begin{aligned}
\mathbb{P} \left( \left| \Delta_{n,2}^k(\theta) \right| \geq \beta \right) &\leq \frac{K}{\beta \sqrt{n}} \sum_{t=1}^n \left( \mathbb{E} \left[ (\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta))^2 \right] \right)^{1/2} \\
&\leq \frac{K}{\beta \sqrt{n}} \sum_{t=1}^n \left( \sum_{i \geq 0} \sum_{j \geq 0} (\lambda_i^t(\theta) - \lambda_i(\theta)) (\lambda_j^t(\theta) - \lambda_j(\theta)) \mathbb{E} [\epsilon_{t-i} \epsilon_{t-j}] \right)^{1/2} \\
&\leq \frac{\sigma_\epsilon K}{\beta \sqrt{n}} \sum_{t=1}^n \left( \sum_{i \geq 0} (\lambda_i^t(\theta) - \lambda_i(\theta))^2 \right)^{1/2} \\
&\leq \frac{\sigma_\epsilon K}{\beta \sqrt{n}} \sum_{t=1}^n \|\lambda(\theta) - \lambda^t(\theta)\|_{\ell^2}.
\end{aligned}$$

We use Lemma 1, the fact that  $d > d_0$  and the fractional version of Cesàro's Lemma<sup>2</sup>, we obtain

$$\mathbb{P} \left( \left| \Delta_{n,2}^k(\theta) \right| \geq \beta \right) \leq \frac{\sigma_\epsilon K}{\beta} \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{1}{t^{1/2+(d-d_0)}} \xrightarrow{n \rightarrow \infty} 0.$$

This proves the expected convergence in probability.

◇ **Step 4: Convergence of  $(\Delta_{n,1}^k(\theta) + \Delta_{n,3}^k(\theta))_n$**

Note now that, for all  $n \geq 1$ , we have

$$\Delta_{n,1}^k(\theta) + \Delta_{n,3}^k(\theta) = \frac{2}{\sqrt{n}} \sum_{t=1}^n \left\{ (\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)) - (\epsilon_t(\theta_0) - \tilde{\epsilon}_t(\theta_0)) \right\} \frac{\partial}{\partial \theta_k} \tilde{\epsilon}_t(\theta_0).$$

By the mean value theorem, there exists  $0 < c_\omega < 1$  such that

$$\left| (\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)) - (\epsilon_t(\theta_0) - \tilde{\epsilon}_t(\theta_0)) \right| \leq \left\| \frac{\partial(\epsilon_t - \tilde{\epsilon}_t)}{\partial \theta} ((1 - c_\omega)\theta + c_\omega \theta_0) \right\|_{\mathbb{R}^{p+q+1}} \|\theta - \theta_0\|_{\mathbb{R}^{p+q+1}}. \quad (49)$$

---

<sup>2</sup>Recall that the fractional version of Cesàro's Lemma states that for  $(h_t)_t$  a sequence of positive real numbers,  $\kappa > 0$  and  $c \geq 0$  we have

$$\lim_{t \rightarrow \infty} h_t t^{1-\kappa} = |\kappa| c \Rightarrow \lim_{n \rightarrow \infty} \frac{1}{n^\kappa} \sum_{t=0}^n h_t = c.$$

Following the same method than in the previous step we obtain

$$\begin{aligned}
\mathbb{E} \left( (\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)) - (\epsilon_t(\theta_0) - \tilde{\epsilon}_t(\theta_0)) \right)^2 &\leq \|\theta - \theta_0\|_{\mathbb{R}^{p+q+1}}^2 \sum_{k=1}^{p+q+1} \mathbb{E} \left[ \left| \frac{\partial(\epsilon_t - \tilde{\epsilon}_t)}{\partial \theta_k} ((1 - c_\omega)\theta + c_\omega \theta_0) \right|^2 \right] \\
&\leq \|\theta - \theta_0\|_{\mathbb{R}^{p+q+1}}^2 \sum_{k=1}^{p+q+1} \sup_{\theta} \mathbb{E} \left[ \left| \frac{\partial(\epsilon_t - \tilde{\epsilon}_t)}{\partial \theta_k}(\theta) \right|^2 \right] \\
&\leq \|\theta - \theta_0\|_{\mathbb{R}^{p+q+1}}^2 \sum_{k=1}^{p+q+1} \sigma_\epsilon^2 \sup_{\theta} \left\| (\dot{\lambda}_k - \dot{\lambda}_k^t)(\theta) \right\|_{\ell^2}^2 \\
&\leq K \|\theta - \theta_0\|_{\mathbb{R}^{p+q+1}}^2 \sup_{d; d_0 \leq d \leq d_2} \left( \frac{1}{t^{1/2+(d-d_0)}} \right)^2 \\
&\leq K \|\theta - \theta_0\|_{\mathbb{R}^{p+q+1}}^2 \frac{1}{t},
\end{aligned} \tag{50}$$

where we have used the fact that the function

$$\theta \mapsto \mathbb{E} \left[ \left| \frac{\partial(\epsilon_t - \tilde{\epsilon}_t)}{\partial \theta_k}(\theta) \right|^2 \right]$$

is bounded and continuous. By (50) and (48), it follows that

$$\mathbb{P} \left( \left| \Delta_{n,1}^k(\theta) + \Delta_{n,3}^k(\theta) \right| \geq \beta \right) \leq \frac{K}{\beta} \|\theta - \theta_0\|_{\mathbb{R}^{p+q+1}} \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{1}{t^{1/2}}$$

and the fractional version of Cesàro's Lemma implies

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \left| \Delta_{n,1}^k(\theta) + \Delta_{n,3}^k(\theta) \right| \geq \beta \right) \leq \frac{K}{\beta} \|\theta - \theta_0\|_{\mathbb{R}^{p+q+1}}. \tag{51}$$

◇ **Step 5: end of the proof** For any  $\varepsilon > 0$ , we choose  $\theta$  such that  $\frac{K}{\beta} \|\theta - \theta_0\|_{\mathbb{R}^{p+q+1}} \leq \varepsilon$ . Then, from (51), there exists  $n_0$  such that for all  $n \geq n_0$ ,

$$\mathbb{P} \left( \left| \Delta_{n,1}^k(\theta) + \Delta_{n,3}^k(\theta) \right| \geq \beta \right) \leq \varepsilon.$$

By Step 2 and 3, one also has for  $n \geq n_0$

$$\mathbb{P} \left( \left| \Delta_{n,2}^k(\theta) + \Delta_{n,4}^k(\theta_0) \right| \geq \beta \right) \leq \varepsilon.$$

Therefore, for all  $n \geq n_0$ ,

$$\begin{aligned}
\mathbb{P} \left( \left| \sqrt{n} \frac{\partial}{\partial \theta_k} Q_n(\theta_0) - \sqrt{n} \frac{\partial}{\partial \theta_k} O_n(\theta_0) \right| \geq 2\beta \right) &\leq \mathbb{P} \left( \left| \Delta_{n,1}^k(\theta) + \Delta_{n,3}^k(\theta) \right| \geq \beta \right) \\
&\quad + \mathbb{P} \left( \left| \Delta_{n,2}^k(\theta) + \Delta_{n,4}^k(\theta_0) \right| \geq \beta \right) \\
&\leq \varepsilon
\end{aligned}$$

and the expected convergence in probability is proved.  $\square$

We show in the following lemma the existence and invertibility of  $J(\theta_0)$ .

**Lemma 6.** *Under Assumptions of Theorem 2, the matrix*

$$J(\theta_0) = \lim_{n \rightarrow \infty} \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} O_n(\theta_0) \right]$$

*exists almost surely and is invertible.*

*Proof.* For all  $1 \leq i, j \leq p + q + 1$ , we have

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} O_n(\theta_0) = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \theta_i \partial \theta_j} \epsilon_t^2(\theta_0) = \frac{2}{n} \sum_{t=1}^n \left\{ \frac{\partial}{\partial \theta_i} \epsilon_t(\theta_0) \frac{\partial}{\partial \theta_j} \epsilon_t(\theta_0) + \epsilon_t(\theta_0) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \epsilon_t(\theta_0) \right\}.$$

Note that in view of (25), (26) and (27), the first and second order derivatives of  $\epsilon_t(\cdot)$  belong to  $\mathbb{L}^2$ . By using the ergodicity of  $(\epsilon_t)_{t \in \mathbb{Z}}$  assumed in Assumption (A2), we deduce that

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} O_n(\theta_0) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 2\mathbb{E} \left[ \frac{\partial}{\partial \theta_i} \epsilon_t(\theta_0) \frac{\partial}{\partial \theta_j} \epsilon_t(\theta_0) \right] + 2\mathbb{E} \left[ \epsilon_t(\theta_0) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \epsilon_t(\theta_0) \right].$$

By (30) and (33),  $\epsilon_t$  and  $\partial \epsilon_t(\theta_0)/\partial \theta$  are non correlated as well as  $\epsilon_t$  and  $\partial^2 \epsilon_t(\theta_0)/\partial \theta \partial \theta$ . Thus we have

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} O_n(\theta_0) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} J(\theta_0)(i, j) := 2\mathbb{E} \left[ \frac{\partial}{\partial \theta_i} \epsilon_t(\theta_0) \frac{\partial}{\partial \theta_j} \epsilon_t(\theta_0) \right]. \quad (52)$$

From (30) and (39) we obtain that

$$\begin{aligned} \mathbb{E} \left[ \frac{\partial}{\partial \theta_i} \epsilon_t(\theta_0) \frac{\partial}{\partial \theta_j} \epsilon_t(\theta_0) \right] &= \mathbb{E} \left[ \left( \sum_{k_1 \geq 1} \dot{\lambda}_{k_1, i}(\theta_0) \epsilon_{t-k_1} \right) \left( \sum_{k_2 \geq 1} \dot{\lambda}_{k_2, j}(\theta_0) \epsilon_{t-k_2} \right) \right] \\ &= \sum_{k_1 \geq 1} \sum_{k_2 \geq 1} \dot{\lambda}_{k_1, i}(\theta_0) \dot{\lambda}_{k_2, j}(\theta_0) \mathbb{E} [\epsilon_{t-k_1} \epsilon_{t-k_2}] \\ &\leq K \sigma_\epsilon^2 \sum_{k_1 \geq 1} \left( \frac{1}{k_1} \right)^2 < \infty. \end{aligned}$$

Therefore  $J(\theta_0)$  exists almost surely.

If the matrix  $J(\theta_0)$  is not invertible, there exists some real constants  $c_1, \dots, c_{p+q+1}$  not all equal to zero such that  $\mathbf{c}' J(\theta_0) \mathbf{c} = \sum_{i=1}^{p+q+1} \sum_{j=1}^{p+q+1} c_j J(\theta_0)(j, i) c_i = 0$ , where  $\mathbf{c} = (c_1, \dots, c_{p+q+1})'$ . In view of (52) we obtain that

$$\sum_{i=1}^{p+q+1} \sum_{j=1}^{p+q+1} \mathbb{E} \left[ \left( c_j \frac{\partial \epsilon_t(\theta_0)}{\partial \theta_j} \right) \left( c_i \frac{\partial \epsilon_t(\theta_0)}{\partial \theta_i} \right) \right] = \mathbb{E} \left[ \left( \sum_{k=1}^{p+q+1} c_k \frac{\partial \epsilon_t(\theta_0)}{\partial \theta_k} \right)^2 \right] = 0,$$

which implies that

$$\sum_{k=1}^{p+q+1} c_k \frac{\partial \epsilon_t(\theta_0)}{\partial \theta_k} = 0 \text{ a.s. or equivalently } \mathbf{c}' \frac{\partial \epsilon_t(\theta_0)}{\partial \theta} = 0 \text{ a.s.} \quad (53)$$

Differentiating the equation (1), we obtain that

$$\mathbf{c}' \frac{\partial}{\partial \theta} \left\{ a_{\theta_0}(L)(1-L)^{d_0} \right\} X_t = \mathbf{c}' \left\{ \frac{\partial}{\partial \theta} b_{\theta_0}(L) \right\} \epsilon_t + b_{\theta_0}(L) \mathbf{c}' \frac{\partial}{\partial \theta} \epsilon_t(\theta_0).$$

and by (53) we may write that

$$\mathbf{c}' \left( \frac{\partial}{\partial \theta} \left\{ a_{\theta_0}(L)(1-L)^{d_0} \right\} X_t - \left\{ \frac{\partial}{\partial \theta} b_{\theta_0}(L) \right\} \epsilon_t \right) = 0 \quad \text{a.s.}$$

It follows that (1) can therefore be rewritten in the form:

$$\left( a_{\theta_0}(L)(1-L)^{d_0} + \mathbf{c}' \frac{\partial}{\partial \theta} \left\{ a_{\theta_0}(L)(1-L)^{d_0} \right\} \right) X_t = \left( b_{\theta_0}(L) + \mathbf{c}' \frac{\partial}{\partial \theta} b_{\theta_0}(L) \right) \epsilon_t, \quad \text{a.s.}$$

Under Assumption (A1) the representation in (1) is unique (see [Hos81]) so

$$\mathbf{c}' \frac{\partial}{\partial \theta} \left\{ a_{\theta_0}(L)(1-L)^{d_0} \right\} = 0 \quad \text{and} \quad (54)$$

$$\mathbf{c}' \frac{\partial}{\partial \theta} b_{\theta_0}(L) = 0. \quad (55)$$

First, (55) implies that

$$\sum_{k=p+1}^{p+q} c_k \frac{\partial}{\partial \theta_k} b_{\theta_0}(L) = \sum_{k=p+1}^{p+q} -c_k L^k = 0$$

and thus  $c_k = 0$  for  $p+1 \leq k \leq p+q$ .

Similarly, (54) yields that

$$\sum_{k=1}^p c_k \frac{\partial}{\partial \theta_k} a_{\theta_0}(L)(1-L)^{d_0} + c_{p+q+1} a_{\theta_0}(L) \frac{\partial(1-L)^d}{\partial d}(d_0) = 0.$$

Since  $\partial(1-L)^d/\partial d = (1-L)^d \ln(1-L)$ , it follows that

$$-\sum_{k=1}^p c_k L^k + c_{p+q+1} \sum_{k \geq 0} e_k L^k = 0,$$

where the sequence  $(e_k)_{k \geq 1}$  is given by the coefficients of the power series  $a_{\theta_0}(L) \ln(1-L)$ . Since  $e_0 = 0$  and  $e_1 = -1$ , we obtain that

$$\begin{aligned} c_1 &= -c_{p+q+1} \\ c_k &= e_k c_{p+q+1} \quad \text{for } k = 2, \dots, p \\ 0 &= e_k c_{p+q+1} \quad \text{for } k \geq p+1. \end{aligned}$$

Since the polynomial  $a_{\theta_0}$  is not the null polynomial, this implies that  $c_{p+q+1} = 0$  and then  $c_k$  for  $1 \leq k \leq p$ . Thus  $\mathbf{c} = 0$  which leads us to a contradiction. Hence  $J(\theta_0)$  is invertible.  $\square$

**Lemma 7.** For any  $1 \leq i, j \leq p+q+1$  and under the assumptions of Theorem 1, we have almost surely

$$\lim_{t \rightarrow \infty} \sup_{\theta \in \Theta_\delta} \left| \frac{\partial}{\partial \theta_i} \epsilon_t(\theta) - \frac{\partial}{\partial \theta_i} \tilde{\epsilon}_t(\theta) \right| = 0 \quad \text{and} \quad \lim_{t \rightarrow \infty} \sup_{\theta \in \Theta_\delta} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \epsilon_t(\theta) - \frac{\partial^2}{\partial \theta_i \partial \theta_j} \tilde{\epsilon}_t(\theta) \right| = 0. \quad (56)$$

*Proof.* The proof uses the same arguments that the proof of Lemma 4 so it is omitted.  $\square$

**Lemma 8.** For any  $1 \leq i, j \leq p + q + 1$  and under the assumptions of Theorem 1, we have almost surely

$$\lim_{n \rightarrow \infty} \left\{ \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\theta_{n,i,j}^*) \right] - J(\theta_0) \right\} = 0 \quad (57)$$

where  $\theta_{n,i,j}^*$  is defined in (41).

*Proof.* For any  $\theta \in \Theta_\delta$ , let

$$J_n(\theta) = \frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta) = \frac{2}{n} \sum_{t=1}^n \left\{ \frac{\partial}{\partial \theta} \tilde{\epsilon}_t(\theta) \right\} \left\{ \frac{\partial}{\partial \theta'} \tilde{\epsilon}_t(\theta) \right\} + \frac{2}{n} \sum_{t=1}^n \tilde{\epsilon}_t(\theta) \frac{\partial^2}{\partial \theta \partial \theta'} \tilde{\epsilon}_t(\theta),$$

and

$$J_n^*(\theta) = \frac{\partial^2}{\partial \theta \partial \theta'} O_n(\theta) = \frac{2}{n} \sum_{t=1}^n \left\{ \frac{\partial}{\partial \theta} \epsilon_t(\theta) \right\} \left\{ \frac{\partial}{\partial \theta'} \epsilon_t(\theta) \right\} + \frac{2}{n} \sum_{t=1}^n \epsilon_t(\theta) \frac{\partial^2}{\partial \theta \partial \theta'} \epsilon_t(\theta).$$

We have

$$\begin{aligned} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\theta_{n,i,j}^*) - J(\theta_0)(i, j) \right| &\leq |J_n(\theta_{n,i,j}^*)(i, j) - J_n^*(\theta_{n,i,j}^*)(i, j)| \\ &\quad + |J_n^*(\theta_{n,i,j}^*)(i, j) - J_n^*(\theta_0)(i, j)| + |J_n^*(\theta_0)(i, j) - J(\theta_0)(i, j)|. \end{aligned} \quad (58)$$

So it is enough to show that the three terms in the right hand side of (58) tend almost-surely to 0 when  $n$  tends to infinity. Following the same arguments as the proof of Lemma 6 and applying the ergodic theorem, we obtain that

$$J_n^*(\theta_0) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 2\mathbb{E} \left[ \frac{\partial}{\partial \theta} \epsilon_t(\theta_0) \frac{\partial}{\partial \theta'} \epsilon_t(\theta_0) \right] = J(\theta_0).$$

Let us now show that the term  $|J_n^*(\theta_{n,i,j}^*)(i, j) - J_n^*(\theta_0)(i, j)|$  converges almost-surely to 0. In view

of (25) and (27), we have

$$\begin{aligned}
\sup_{\theta \in \Theta_\delta} \left\| \frac{\partial}{\partial \theta} \left( \frac{\partial}{\partial \theta_i} \epsilon_t(\theta) \frac{\partial}{\partial \theta_j} \epsilon_t(\theta) \right) \right\| &= \sup_{\theta \in \Theta_\delta} \left\| \frac{\partial}{\partial \theta} \left\{ \left( \sum_{k_1 \geq 1} \frac{\partial}{\partial \theta_i} \gamma_{k_1}(\theta) X_{t-k_1} \right) \left( \sum_{k_2 \geq 1} \frac{\partial}{\partial \theta_j} \gamma_{k_2}(\theta) X_{t-k_2} \right) \right\} \right\| \\
&= \sup_{\theta \in \Theta_\delta} \left\| \frac{\partial}{\partial \theta} \left( \sum_{k_1, k_2 \geq 1} \frac{\partial}{\partial \theta_i} \gamma_{k_1}(\theta) \frac{\partial}{\partial \theta_j} \gamma_{k_2}(\theta) X_{t-k_1} X_{t-k_2} \right) \right\| \\
&\leq \sup_{\theta \in \Theta_\delta} \left\| \sum_{k_1, k_2 \geq 1} \left( \frac{\partial}{\partial \theta} \frac{\partial}{\partial \theta_i} \gamma_{k_1}(\theta) \right) \frac{\partial}{\partial \theta_j} \gamma_{k_2}(\theta) X_{t-k_1} X_{t-k_2} \right\| \\
&\quad + \sup_{\theta \in \Theta_\delta} \left\| \sum_{k_1, k_2 \geq 1} \frac{\partial}{\partial \theta_i} \gamma_{k_1}(\theta) \left( \frac{\partial}{\partial \theta} \frac{\partial}{\partial \theta_j} \gamma_{k_2}(\theta) \right) X_{t-k_1} X_{t-k_2} \right\| \\
&\leq \sum_{k_1, k_2 \geq 1} \sup_{\theta \in \Theta_\delta} \left\| \frac{\partial}{\partial \theta} \frac{\partial}{\partial \theta_i} \gamma_{k_1}(\theta) \right\| \sup_{\theta \in \Theta_\delta} \left\| \frac{\partial}{\partial \theta_j} \gamma_{k_2}(\theta) \right\| |X_{t-k_1}| |X_{t-k_2}| \\
&\quad + \sum_{k_1, k_2 \geq 1} \sup_{\theta \in \Theta_\delta} \left\| \frac{\partial}{\partial \theta_i} \gamma_{k_1}(\theta) \right\| \sup_{\theta \in \Theta_\delta} \left\| \frac{\partial}{\partial \theta} \frac{\partial}{\partial \theta_j} \gamma_{k_2}(\theta) \right\| |X_{t-k_1}| |X_{t-k_2}| \\
&\leq K \sum_{k_1, k_2 \geq 1} (\log(k_1))^2 k_1^{-1-d_1} \log(k_2) k_2^{-1-d_1} |X_{t-k_1}| |X_{t-k_2}| \\
&\quad + K \sum_{k_1, k_2 \geq 1} \log(k_1) k_1^{-1-d_1} (\log(k_2))^2 k_2^{-1-d_1} |X_{t-k_1}| |X_{t-k_2}|.
\end{aligned}$$

Consequently we obtain

$$\begin{aligned}
\mathbb{E}_{\theta_0} \left[ \sup_{\theta \in \Theta_\delta} \left\| \frac{\partial}{\partial \theta} \left( \frac{\partial}{\partial \theta_i} \epsilon_t(\theta) \frac{\partial}{\partial \theta_j} \epsilon_t(\theta) \right) \right\| \right] &\leq K \sum_{k_1, k_2 \geq 1} (\log(k_1))^2 k_1^{-1-d_1} \log(k_2) k_2^{-1-d_1} \sup_{t \in \mathbb{Z}} \mathbb{E}_{\theta_0} |X_t|^2 \\
&\quad + K \sum_{k_1, k_2 \geq 1} \log(k_1) k_1^{-1-d_1} (\log(k_2))^2 k_2^{-1-d_1} \sup_{t \in \mathbb{Z}} \mathbb{E}_{\theta_0} |X_t|^2 \\
&\leq K.
\end{aligned} \tag{59}$$

Following the same approach used in (59), we have

$$\mathbb{E}_{\theta_0} \left[ \sup_{\theta \in \Theta_\delta} \left\| \frac{\partial}{\partial \theta} \left\{ \epsilon_t(\theta) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \epsilon_t(\theta) \right\} \right\| \right] < \infty. \tag{60}$$

A Taylor expansion implies that there exists a random variable  $\theta_{n,i,j}^{**}$ 's between  $\theta_{n,i,j}^*$  and  $\theta_0$  such

that

$$\begin{aligned}
 |J_n^*(\theta_{n,i,j}^*)(i,j) - J_n^*(\theta_0)(i,j)| &= \left| \frac{\partial}{\partial \theta} J_n^*(\theta_{n,i,j}^{**})(i,j) \cdot (\theta_{n,i,j}^* - \theta_0) \right| \\
 &\leq \sup_{\theta \in \Theta_\delta} \left\| \frac{\partial}{\partial \theta} J_n^*(\theta)(i,j) \right\| \|\theta_{n,i,j}^* - \theta_0\| \\
 &\leq \frac{2}{n} \sum_{t=1}^n \sup_{\theta \in \Theta_\delta} \left\| \frac{\partial}{\partial \theta} \left( \frac{\partial}{\partial \theta_i} \epsilon_t(\theta) \frac{\partial}{\partial \theta_j} \epsilon_t(\theta) \right) \right\| \|\theta_{n,i,j}^* - \theta_0\| \\
 &\quad + \frac{2}{n} \sum_{t=1}^n \sup_{\theta \in \Theta_\delta} \left\| \frac{\partial}{\partial \theta} \left\{ \epsilon_t(\theta) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \epsilon_t(\theta) \right\} \right\| \|\theta_{n,i,j}^* - \theta_0\|.
 \end{aligned}$$

By Theorem 1, the ergodic theorem, (59) and (60) imply that  $\lim_{n \rightarrow \infty} |J_n^*(\theta_{n,i,j}^*)(i,j) - J_n^*(\theta_0)(i,j)| = 0$  a.s.

To prove the almost-sure convergence of the first term of the right hand side of (58) it suffices to show that

$$\frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \Theta_\delta} \left| \frac{\partial}{\partial \theta_i} \epsilon_t(\theta) \frac{\partial}{\partial \theta_j} \epsilon_t(\theta) - \frac{\partial}{\partial \theta_i} \tilde{\epsilon}_t(\theta) \frac{\partial}{\partial \theta_j} \tilde{\epsilon}_t(\theta) \right|$$

and

$$\frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \Theta_\delta} \left| \epsilon_t(\theta) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \epsilon_t(\theta) - \tilde{\epsilon}_t(\theta) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \tilde{\epsilon}_t(\theta) \right|$$

converge almost-surely to 0. On one hand we have

$$\begin{aligned}
 &\frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \Theta_\delta} \left| \frac{\partial}{\partial \theta_i} \epsilon_t(\theta) \frac{\partial}{\partial \theta_j} \epsilon_t(\theta) - \frac{\partial}{\partial \theta_i} \tilde{\epsilon}_t(\theta) \frac{\partial}{\partial \theta_j} \tilde{\epsilon}_t(\theta) \right| \\
 &\leq \frac{1}{n} \sum_{t=1}^n \left\{ \sup_{\theta \in \Theta_\delta} \left| \frac{\partial}{\partial \theta_i} \epsilon_t(\theta) - \frac{\partial}{\partial \theta_i} \tilde{\epsilon}_t(\theta) \right| \sup_{\theta \in \Theta_\delta} \left| \frac{\partial}{\partial \theta_j} \epsilon_t(\theta) \right| + \sup_{\theta \in \Theta_\delta} \left| \frac{\partial}{\partial \theta_i} \tilde{\epsilon}_t(\theta) \right| \sup_{\theta \in \Theta_\delta} \left| \frac{\partial}{\partial \theta_j} \tilde{\epsilon}_t(\theta) - \frac{\partial}{\partial \theta_j} \epsilon_t(\theta) \right| \right\} \\
 &\leq \left( \frac{1}{n} \sum_{t=1}^n \left( \sup_{\theta \in \Theta_\delta} \left| \frac{\partial}{\partial \theta_i} \epsilon_t(\theta) - \frac{\partial}{\partial \theta_i} \tilde{\epsilon}_t(\theta) \right| \right)^2 \right)^{1/2} \left( \frac{1}{n} \sum_{t=1}^n \left( \sup_{\theta \in \Theta_\delta} \left| \frac{\partial}{\partial \theta_j} \epsilon_t(\theta) \right| \right)^2 \right)^{1/2} \\
 &\quad + \left( \frac{1}{n} \sum_{t=1}^n \left( \sup_{\theta \in \Theta_\delta} \left| \frac{\partial}{\partial \theta_i} \tilde{\epsilon}_t(\theta) \right| \right)^2 \right)^{1/2} \left( \frac{1}{n} \sum_{t=1}^n \left( \sup_{\theta \in \Theta_\delta} \left| \frac{\partial}{\partial \theta_j} \tilde{\epsilon}_t(\theta) - \frac{\partial}{\partial \theta_j} \epsilon_t(\theta) \right| \right)^2 \right)^{1/2}.
 \end{aligned}$$

In view of (25) and (27) it follows that

$$\mathbb{E}_{\theta_0} \left[ \left( \sup_{\theta \in \Theta_\delta} \left| \frac{\partial}{\partial \theta_j} \epsilon_t(\theta) \right| \right)^2 \right] \leq \sup_{t \in \mathbb{Z}} \mathbb{E}_{\theta_0} |X_t|^2 \left( \sum_{k_1 \geq 1} \log(k_1) k_1^{-1-d_1} \right)^2 < \infty.$$

Similar commutations imply that

$$\mathbb{E}_{\theta_0} \left[ \left( \sup_{\theta \in \Theta_\delta} \left| \frac{\partial}{\partial \theta_j} \tilde{\epsilon}_t(\theta) \right| \right)^2 \right] < \infty.$$



Cesàro's Lemma, Lemma (7) and the ergodic theorem yield

$$\frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \Theta_\delta} \left| \frac{\partial}{\partial \theta_i} \epsilon_t(\theta) \frac{\partial}{\partial \theta_j} \epsilon_t(\theta) - \frac{\partial}{\partial \theta_i} \tilde{\epsilon}_t(\theta) \frac{\partial}{\partial \theta_j} \tilde{\epsilon}_t(\theta) \right| \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0.$$

On the other hand, one similarly may prove that

$$\frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \Theta_\delta} \left| \epsilon_t(\theta) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \epsilon_t(\theta) - \tilde{\epsilon}_t(\theta) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \tilde{\epsilon}_t(\theta) \right| \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0.$$

Thus

$$\sup_{\theta \in \Theta_\delta} \|J_n(\theta) - J_n^*(\theta)\| \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0$$

and the lemma is proved.  $\square$

The following lemma states the existence of the matrix  $I(\theta_0)$ .

**Lemma 9.** *Under the assumptions of Theorem 2, the matrix*

$$I(\theta_0) = \lim_{n \rightarrow \infty} \text{Var} \left\{ \sqrt{n} \frac{\partial}{\partial \theta} O_n(\theta_0) \right\}$$

exists.

*Proof.* By the stationarity of  $(H_t(\theta_0))_{t \in \mathbb{Z}}$  (remind that this process is defined in (7)), we have

$$\begin{aligned} \text{Var} \left\{ \sqrt{n} \frac{\partial}{\partial \theta} O_n(\theta_0) \right\} &= \text{Var} \left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^n H_t(\theta_0) \right\} \\ &= \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n \text{Cov} \{H_t(\theta_0), H_s(\theta_0)\} \\ &= \frac{1}{n} \sum_{h=-n+1}^{n-1} (n - |h|) \text{Cov} \{H_t(\theta_0), H_{t-h}(\theta_0)\}. \end{aligned}$$

By the dominated convergence theorem, the matrix  $I(\theta_0)$  exists and is given by

$$I(\theta_0) = \sum_{h=-\infty}^{\infty} \text{Cov} \{H_t(\theta_0), H_{t-h}(\theta_0)\}$$

whenever

$$\sum_{h=-\infty}^{\infty} \|\text{Cov} \{H_t(\theta_0), H_{t-h}(\theta_0)\}\| < \infty. \quad (61)$$

For  $s \in \mathbb{Z}$  and  $1 \leq k \leq p + q + 1$ , we denote  $H_{s,k}(\theta_0) = 2\epsilon_s(\theta_0) \frac{\partial}{\partial \theta_k} \epsilon_s(\theta_0)$  the  $k$ -th entry of

$H_s(\theta_0)$ . In view of (30) we have

$$\begin{aligned} |Cov\{H_{t,i}(\theta_0), H_{t-h,j}(\theta_0)\}| &= 4 \left| Cov \left( \sum_{k_1 \geq 1} \dot{\lambda}_{k_1,i}(\theta_0) \epsilon_t \epsilon_{t-k_1}, \sum_{k_2 \geq 1} \dot{\lambda}_{k_2,j}(\theta_0) \epsilon_{t-h} \epsilon_{t-h-k_2} \right) \right| \\ &\leq 4 \sum_{k_1 \geq 1} \sum_{k_2 \geq 1} \left| \dot{\lambda}_{k_1,i}(\theta_0) \right| \left| \dot{\lambda}_{k_2,j}(\theta_0) \right| |\mathbb{E}[\epsilon_t \epsilon_{t-k_1} \epsilon_{t-h} \epsilon_{t-h-k_2}]| \\ &\leq \sum_{k_1 \geq 1} \sum_{k_2 \geq 1} \frac{K}{k_1 k_2} |\mathbb{E}[\epsilon_t \epsilon_{t-k_1} \epsilon_{t-h} \epsilon_{t-h-k_2}]| \end{aligned}$$

where we have used Lemma 3. It follows that

$$\begin{aligned} \sum_{h=-\infty}^{\infty} |Cov\{H_{t,i}(\theta_0), H_{t-h,j}(\theta_0)\}| &\leq \sum_{h \in \mathbb{Z} \setminus \{0\}} \sum_{k_1 \geq 1} \sum_{k_2 \geq 1} \frac{K}{k_1 k_2} |\text{cum}(\epsilon_t, \epsilon_{t-k_1}, \epsilon_{t-h}, \epsilon_{t-h-k_2})| \\ &\quad + \sum_{k_1 \geq 1} \sum_{k_2 \geq 1} \frac{K}{k_1 k_2} |\mathbb{E}[\epsilon_t \epsilon_{t-k_1} \epsilon_t \epsilon_{t-k_2}]|. \end{aligned}$$

Thanks to the stationarity of  $(\epsilon_t)_{t \in \mathbb{Z}}$  and Assumption (A4') with  $\tau = 4$  we deduce that

$$\begin{aligned} \sum_{h=-\infty}^{\infty} |Cov\{H_{t,i}(\theta_0), H_{t-h,j}(\theta_0)\}| &\leq \sum_{h \in \mathbb{Z} \setminus \{0\}} \sum_{k_1 \geq 1} \sum_{k_2 \geq 1} \frac{K}{k_1 k_2} |\text{cum}(\epsilon_0, \epsilon_{-k_1}, \epsilon_{-h}, \epsilon_{-h-k_2})| \\ &\quad + \sum_{k_1 \geq 1} \sum_{k_2 \geq 1} \frac{K}{k_1 k_2} |\mathbb{E}[\epsilon_0 \epsilon_{-k_1} \epsilon_0 \epsilon_{-k_2}]| \\ &\leq K \sum_{h,k,l \in \mathbb{Z}} |\text{cum}(\epsilon_0, \epsilon_k, \epsilon_h, \epsilon_l)| \\ &\quad + \sum_{k_1 \geq 1} \sum_{k_2 \geq 1} \frac{K}{k_1 k_2} \left( |\text{cum}(\epsilon_0, \epsilon_{-k_1}, \epsilon_0, \epsilon_{-k_2})| \right. \\ &\quad \left. + \sigma_\epsilon^2 |\mathbb{E}[\epsilon_{-k_1} \epsilon_{-k_2}]| \right) \\ &\leq K \sum_{h,k,l \in \mathbb{Z}} |\text{cum}(\epsilon_0, \epsilon_k, \epsilon_h, \epsilon_l)| + K \sigma_\epsilon^4 \sum_{k_1 \geq 1} \left( \frac{1}{k_1} \right)^2 \leq K \end{aligned}$$

and we obtain the expected result.  $\square$

**Lemma 10.** Under Assumptions of Theorem 2, the random vector  $\sqrt{n}(\partial/\partial\theta)O_n(\theta_0)$  has a limiting normal distribution with mean 0 and covariance matrix  $I(\theta_0)$ .

*Proof.* Observe that for any  $t \in \mathbb{Z}$

$$\mathbb{E} \left[ \epsilon_t \frac{\partial}{\partial \theta} \epsilon_t(\theta_0) \right] = 0 \quad (62)$$

because  $\partial \epsilon_t(\theta_0)/\partial \theta$  belongs to the Hilbert space  $\mathbf{H}_\epsilon(t-1)$ , linearly generated by the family  $(\epsilon_s)_{s \leq t-1}$ . Therefore we have

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[ \sqrt{n} \frac{\partial}{\partial \theta} O_n(\theta_0) \right] = \lim_{n \rightarrow \infty} \frac{2}{\sqrt{n}} \sum_{t=1}^n \mathbb{E} \left[ \epsilon_t \frac{\partial}{\partial \theta} \epsilon_t(\theta_0) \right] = 0.$$

For  $i \geq 1$ , we denote  $\Lambda_i(\theta_0) = (\dot{\lambda}_{i,1}(\theta_0), \dots, \dot{\lambda}_{i,p+q+1}(\theta_0))'$  and we introduce for  $r \geq 1$

$$H_{t,r}(\theta_0) = 2 \sum_{i=1}^r \Lambda_i(\theta_0) \epsilon_t \epsilon_{t-i} \quad \text{and} \quad G_{t,r}(\theta_0) = 2 \sum_{i \geq r+1} \Lambda_i(\theta_0) \epsilon_t \epsilon_{t-i}.$$

From (30) we have

$$\sqrt{n} \frac{\partial}{\partial \theta} O_n(\theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^n H_{t,r}(\theta_0) + \frac{1}{\sqrt{n}} \sum_{t=1}^n G_{t,r}(\theta_0).$$

Since  $H_{t,r}(\theta_0)$  is a function of finite number of values of the process  $(\epsilon_t)_{t \in \mathbb{Z}}$ , the stationary process  $(H_{t,r}(\theta_0))_{t \in \mathbb{Z}}$  satisfies a mixing property (see Theorem 14.1 in [Dav94], p. 210) of the form **(A4)**. The central limit theorem for strongly mixing processes (see [Her84]) implies that  $(1/\sqrt{n}) \sum_{t=1}^n H_{t,r}(\theta_0)$  has a limiting  $\mathcal{N}(0, I_r(\theta_0))$  distribution with

$$I_r(\theta_0) = \lim_{n \rightarrow \infty} \text{Var} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n H_{t,r}(\theta_0) \right).$$

Since  $\frac{1}{\sqrt{n}} \sum_{t=1}^n H_{t,r}(\theta_0)$  and  $\frac{1}{\sqrt{n}} \sum_{t=1}^n H_t(\theta_0)$  have zero expectation, we shall have

$$\lim_{r \rightarrow \infty} \text{Var} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n H_{t,r}(\theta_0) \right) = \text{Var} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n H_t(\theta_0) \right) = \text{Var} \left\{ \sqrt{n} \frac{\partial}{\partial \theta} O_n(\theta_0) \right\},$$

as soon as

$$\lim_{r \rightarrow \infty} \mathbb{E} \left[ \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^n H_t(\theta_0) - \frac{1}{\sqrt{n}} \sum_{t=1}^n H_{t,r}(\theta_0) \right\|_{\mathbb{R}^{p+q+1}}^2 \right] = 0. \quad (63)$$

As a consequence we will have  $\lim_{r \rightarrow \infty} I_r(\theta_0) = I(\theta_0)$ . The limit in (63) is obtained as follows:

$$\begin{aligned} \mathbb{E} \left[ \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^n H_t(\theta_0) - \frac{1}{\sqrt{n}} \sum_{t=1}^n H_{t,r}(\theta_0) \right\|_{\mathbb{R}^{p+q+1}}^2 \right] &= \mathbb{E} \left[ \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^n G_{t,r}(\theta_0) \right\|_{\mathbb{R}^{p+q+1}}^2 \right] \\ &\leq \frac{4}{n} \sum_{l=1}^{p+q+1} \mathbb{E} \left[ \left( \sum_{t=1}^n \sum_{k \geq r+1} \dot{\lambda}_{k,l}(\theta_0) \epsilon_{t-k} \epsilon_t \right)^2 \right] \\ &\leq \frac{4}{n} \sum_{l=1}^{p+q+1} \sum_{t=1}^n \sum_{s=1}^n \sum_{k \geq r+1} \sum_{j \geq r+1} \left| \dot{\lambda}_{k,l}(\theta_0) \right| \left| \dot{\lambda}_{j,l}(\theta_0) \right| \mathbb{E} [\epsilon_{t-k} \epsilon_t \epsilon_{s-j} \epsilon_s], \end{aligned}$$

We use successively the stationarity, Lemma 3 and Assumption **(A4')** with  $\tau = 4$  in order to obtain

that

$$\begin{aligned}
& \mathbb{E} \left[ \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^n H_t(\theta_0) - \frac{1}{\sqrt{n}} \sum_{t=1}^n H_{t,r}(\theta_0) \right\|_{\mathbb{R}^{p+q+1}}^2 \right] \\
& \leq \frac{4}{n} \sum_{l=1}^{p+q+1} \sum_{h=1-n}^{n-1} \sum_{k \geq r+1} \sum_{j \geq r+1} \left| \dot{\lambda}_{k,l}(\theta_0) \right| \left| \dot{\lambda}_{j,l}(\theta_0) \right| (n - |h|) |\mathbb{E}[\epsilon_{t-k} \epsilon_t \epsilon_{t-h-j} \epsilon_{t-h}]| \\
& \leq 4 \sum_{l=1}^{p+q+1} \sum_{h=-\infty}^{\infty} \sum_{k \geq r+1} \sum_{j \geq r+1} \left| \dot{\lambda}_{k,l}(\theta_0) \right| \left| \dot{\lambda}_{j,l}(\theta_0) \right| |\mathbb{E}[\epsilon_{t-k} \epsilon_t \epsilon_{t-h-j} \epsilon_{t-h}]| \\
& \leq \frac{K}{(r+1)^2} \sum_{h \neq 0} \sum_{k \geq r+1} \sum_{j=-\infty}^{\infty} |\text{cum}(\epsilon_0, \epsilon_{-k}, \epsilon_{-j}, \epsilon_{-h})| \\
& \quad + \frac{K}{(r+1)^2} \sum_{k \geq r+1} \sum_{j \geq r+1} |\text{cum}(\epsilon_0, \epsilon_{-k}, \epsilon_{-j}, \epsilon_0)| + K \sigma_\epsilon^4 \sum_{k \geq r+1} \left( \frac{1}{k} \right)^2
\end{aligned}$$

and we obtain the convergence stated in (63) when  $r \rightarrow \infty$ .

Using Theorem 7.7.1 and Corollary 7.7.1 of Anderson (see [And71] pages 425-426), the Lemma is proved once we have, uniformly in  $n$ ,

$$\text{Var} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n G_{t,r}(\theta_0) \right) \xrightarrow{r \rightarrow \infty} 0.$$

Arguing as before we may write

$$\begin{aligned}
\left[ \text{Var} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n G_{t,r}(\theta_0) \right) \right]_{ij} &= \left[ \text{Var} \left( \frac{2}{\sqrt{n}} \sum_{t=1}^n \sum_{k \geq r+1} \Lambda_k(\theta_0) \epsilon_{t-k} \epsilon_t \right) \right]_{ij} \\
&= \frac{4}{n} \sum_{t=1}^n \sum_{s=1}^n \sum_{k_1 \geq r+1} \sum_{k_2 \geq r+1} \dot{\lambda}_{k_1,i}(\theta_0) \dot{\lambda}_{k_2,j}(\theta_0) \mathbb{E}[\epsilon_{t-k_1} \epsilon_t \epsilon_{s-k_2} \epsilon_s] \\
&\leq 4 \sum_{h=-\infty}^{\infty} \sum_{k_1, k_2 \geq r+1} \left| \dot{\lambda}_{k_1,i}(\theta_0) \dot{\lambda}_{k_2,j}(\theta_0) \right| |\mathbb{E}[\epsilon_{t-k_1} \epsilon_t \epsilon_{t-h-k_2} \epsilon_{t-h}]|.
\end{aligned}$$

and we obtain that

$$\sup_n \text{Var} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n G_{t,r}(\theta_0) \right) \xrightarrow{r \rightarrow \infty} 0, \quad (64)$$

which completes the proof.  $\square$

No we can end this quite long proof of the asymptotic normality result.

## Proof of Theorem 2

In view of Lemma 5, the equation (42) can be rewritten in the form:

$$o_{\mathbb{P}}(1) = \sqrt{n} \frac{\partial}{\partial \theta} O_n(\theta_0) + \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\theta_{n,i,j}^*) \right] \sqrt{n} (\hat{\theta}_n - \theta_0).$$

From Lemma 10  $\sqrt{n}(\hat{\theta}_n - \theta_0)(\partial^2/\partial \theta_i \partial \theta_j)Q_n(\theta_{n,i,j}^*)$  converges in distribution to  $\mathcal{N}(0, I(\theta_0))$ . Using Lemma 8 and Slutsky's theorem we deduce that

$$\left( \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\theta_{n,i,j}^*) \right], \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\theta_{n,i,j}^*) \right] \sqrt{n}(\hat{\theta}_n - \theta_0) \right)$$

converges in distribution to  $(J(\theta_0), Z)$  with  $\mathbb{P}_Z = \mathcal{N}(0, I)$ . Consider now the function  $h : \mathbb{R}^{(p+q+1) \times (p+q+1)} \times \mathbb{R}^{p+q+1} \rightarrow \mathbb{R}^{p+q+1}$  that maps  $(A, X)$  to  $A^{-1}X$ . If  $D_h$  denotes the set of discontinuity points of  $h$ , we have  $\mathbb{P}((J(\theta_0), Z) \in D_h) = 0$ . By the continuous mapping theorem

$$h \left( \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\theta_{n,i,j}^*) \right], \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n(\theta_{n,i,j}^*) \right] \sqrt{n}(\hat{\theta}_n - \theta_0) \right)$$

converges in distribution to  $h(J(\theta_0), Z)$  and thus  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  has a limiting normal distribution with mean 0 and covariance matrix  $J^{-1}(\theta_0)I(\theta_0)J^{-1}(\theta_0)$ . The proof of Theorem 2 is then completed.

### 6.4. Proof of the convergence of the variance matrix estimator

We show in this section the convergence in probability of  $\hat{\Omega} := \hat{J}_n^{-1} \hat{I}_n^{SP} \hat{J}_n^{-1}$  to  $\Omega$ , which is an adaptation of the arguments used in [BMCF12].

Using the same approach as that followed in Lemma 8, we show that  $\hat{J}_n$  converges almost surely to  $J$ . We give below the proof of the convergence in probability of the estimator  $\hat{I}_n^{SP}$ , obtained using the approach of the spectral density, towards  $I$ .

We recall that the matrix norm used is given by  $\|A\| = \sup_{\|x\| \leq 1} \|Ax\| = \rho^{1/2}(A'A)$ , when  $A$  is a  $\mathbb{R}^{k_1 \times k_2}$  matrix,  $\|x\|^2 = x'x$  is the Euclidean norm of the vector  $x \in \mathbb{R}^{k_2}$ , and  $\rho(\cdot)$  denotes the spectral radius. This norm satisfies

$$\|A\|^2 \leq \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} a_{ij}^2, \quad (65)$$

with  $a_{ij}$  the entries of  $A \in \mathbb{R}^{k_1 \times k_2}$ . The choice of the norm is crucial for the following results to hold (with e.g. the Euclidean norm, this result is not valid).

We denote

$$\Sigma_{H, \underline{H}_r} = \mathbb{E} H_t \underline{H}_{r,t}', \quad \Sigma_H = \mathbb{E} H_t H_t', \quad \Sigma_{\underline{H}_r} = \mathbb{E} \underline{H}_{r,t} \underline{H}_{r,t}'$$

where  $H_t := H_t(\theta_0)$  is defined in (7) and  $\underline{H}_{r,t} = (H_{t-1}', \dots, H_{t-r}')'$ . For any  $n \geq 1$ , we have

$$\begin{aligned} \hat{I}_n^{SP} &= \hat{\Phi}_r^{-1}(1) \hat{\Sigma}_{\hat{\theta}_r} \hat{\Phi}_r'^{-1}(1) \\ &= \left( \hat{\Phi}_r^{-1}(1) - \Phi^{-1}(1) \right) \hat{\Sigma}_{\hat{\theta}_r} \hat{\Phi}_r'^{-1}(1) + \Phi^{-1}(1) \left( \hat{\Sigma}_{\hat{\theta}_r} - \Sigma_u \right) \hat{\Phi}_r'^{-1}(1) \\ &\quad + \Phi^{-1}(1) \Sigma_u \left( \hat{\Phi}_r'^{-1}(1) - \Phi'^{-1}(1) \right) + \Phi^{-1}(1) \Sigma_u \Phi'^{-1}(1). \end{aligned}$$

We then obtain

$$\begin{aligned}
\left\| \hat{I}_n^{\text{SP}} - I(\theta_0) \right\| &\leq \left\| \hat{\Phi}_r^{-1}(1) - \Phi^{-1}(1) \right\| \left\| \hat{\Sigma}_{\hat{u}_r} \right\| \left\| \hat{\Phi}_r'^{-1}(1) \right\| + \left\| \Phi^{-1}(1) \right\| \left\| \hat{\Sigma}_{\hat{u}_r} - \Sigma_u \right\| \left\| \hat{\Phi}_r'^{-1}(1) \right\| \\
&\quad + \left\| \Phi^{-1}(1) \right\| \left\| \Sigma_u \right\| \left\| \hat{\Phi}_r'^{-1}(1) - \Phi'^{-1}(1) \right\| \\
&\leq \left\| \hat{\Phi}_r^{-1}(1) - \Phi^{-1}(1) \right\| \left( \left\| \hat{\Sigma}_{\hat{u}_r} \right\| \left\| \hat{\Phi}_r'^{-1}(1) \right\| + \left\| \Phi^{-1}(1) \right\| \left\| \Sigma_u \right\| \right) \\
&\quad + \left\| \hat{\Sigma}_{\hat{u}_r} - \Sigma_u \right\| \left\| \hat{\Phi}_r'^{-1}(1) \right\| \left\| \Phi^{-1}(1) \right\| \\
&\leq \left\| \hat{\Phi}_r^{-1}(1) \right\| \left\| \Phi(1) - \hat{\Phi}_r(1) \right\| \left\| \Phi^{-1}(1) \right\| \left( \left\| \hat{\Sigma}_{\hat{u}_r} \right\| \left\| \hat{\Phi}_r'^{-1}(1) \right\| + \left\| \Phi^{-1}(1) \right\| \left\| \Sigma_u \right\| \right) \\
&\quad + \left\| \hat{\Sigma}_{\hat{u}_r} - \Sigma_u \right\| \left\| \hat{\Phi}_r'^{-1}(1) \right\| \left\| \Phi^{-1}(1) \right\|. \tag{66}
\end{aligned}$$

In view of (66), to prove the convergence in probability of  $\hat{I}_n^{\text{SP}}$  to  $I(\theta_0)$ , it suffices to show that  $\hat{\Phi}_r(1) \rightarrow \Phi(1)$  and  $\hat{\Sigma}_{\hat{u}_r} \rightarrow \Sigma_u$  in probability. Let the  $r \times 1$  vector  $1_r = (1, \dots, 1)'$  and the  $r(p+q+1) \times (p+q+1)$  matrix  $\mathbf{E}_r = I_{p+q+1} \otimes 1_r$ , where  $\otimes$  denotes the matrix Kronecker product and  $I_m$  the  $m \times m$  identity matrix. Write  $\underline{\Phi}_r^* = (\Phi_1, \dots, \Phi_r)$  where the  $\Phi_i$ 's are defined by (8). We have

$$\begin{aligned}
\left\| \hat{\Phi}_r(1) - \Phi(1) \right\| &= \left\| \sum_{k=1}^r \hat{\Phi}_{r,k} - \sum_{k=1}^r \Phi_{r,k} + \sum_{k=1}^r \Phi_{r,k} - \sum_{k=1}^{\infty} \Phi_k \right\| \\
&\leq \left\| \sum_{k=1}^r (\hat{\Phi}_{r,k} - \Phi_{r,k}) \right\| + \left\| \sum_{k=1}^r (\Phi_{r,k} - \Phi_k) \right\| + \left\| \sum_{k=r+1}^{\infty} \Phi_k \right\| \\
&\leq \left\| (\hat{\Phi}_r - \underline{\Phi}_r) \mathbf{E}_r \right\| + \left\| (\underline{\Phi}_r^* - \underline{\Phi}_r) \mathbf{E}_r \right\| + \left\| \sum_{k=r+1}^{\infty} \Phi_k \right\| \\
&\leq \sqrt{p+q+1} \sqrt{r} \left( \left\| \hat{\Phi}_r - \underline{\Phi}_r \right\| + \left\| \underline{\Phi}_r^* - \underline{\Phi}_r \right\| \right) + \left\| \sum_{k=r+1}^{\infty} \Phi_k \right\|. \tag{67}
\end{aligned}$$

Under the assumptions of Theorem 3 (see (9)) we have

$$\left\| \sum_{k=r+1}^{\infty} \Phi_k \right\| \leq \sum_{k=r+1}^{\infty} \left\| \Phi_k \right\| \leq K \sum_{k=r+1}^{\infty} \rho^k \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0.$$

Therefore it is enough to show that  $\sqrt{r} \left\| \hat{\Phi}_r - \underline{\Phi}_r \right\|$  and  $\sqrt{r} \left\| \underline{\Phi}_r^* - \underline{\Phi}_r \right\|$  converge in probability towards 0 in order to obtain the convergence in probability of  $\hat{\Phi}_r(1)$  towards  $\Phi(1)$ . From (10) we have

$$H_t(\theta_0) = \underline{\Phi}_r \underline{H}_{r,t}(\theta_0) + u_{r,t}, \tag{68}$$

and thus

$$\Sigma_{u_r} = \text{Var}(u_{r,t}) = \mathbb{E} \left[ u_{r,t} (H_t(\theta_0) - \underline{\Phi}_r \underline{H}_{r,t}(\theta_0))' \right].$$

The vector  $u_{r,t}$  is orthogonal to  $\underline{H}_{r,t}(\theta_0)$ . Therefore

$$\begin{aligned}
\text{Var}(u_{r,t}) &= \mathbb{E} \left[ (H_t(\theta_0) - \underline{\Phi}_r \underline{H}_{r,t}(\theta_0)) H_t'(\theta_0) \right] \\
&= \Sigma_H - \underline{\Phi}_r \Sigma_{H, \underline{H}_r}'.
\end{aligned}$$

Consequently the least squares estimator of  $\Sigma_{u_r}$  can be rewritten in the form:

$$\hat{\Sigma}_{\hat{u}_r} = \hat{\Sigma}_{\hat{H}} - \hat{\Phi}_r \hat{\Sigma}'_{\hat{H}, \hat{H}_r}, \quad (69)$$

where

$$\hat{\Sigma}_{\hat{H}} = \frac{1}{n} \sum_{t=1}^n \hat{H}_t \hat{H}_t'. \quad (70)$$

Similar arguments combined with (8) yield

$$\begin{aligned} \Sigma_u &= \mathbb{E} [u_t u_t'] = \mathbb{E} [u_t H_t'(\theta_0)] \\ &= \mathbb{E} [H_t(\theta_0) H_t'(\theta_0)] - \sum_{k=1}^r \Phi_k \mathbb{E} [H_{t-k}(\theta_0) H_t'(\theta_0)] - \sum_{k=r+1}^{\infty} \Phi_k \mathbb{E} [H_{t-k}(\theta_0) H_t'(\theta_0)] \\ &= \Sigma_H - \Phi_r^* \Sigma'_{H, H_r} - \sum_{k=r+1}^{\infty} \Phi_k \mathbb{E} [H_{t-k}(\theta_0) H_t'(\theta_0)]. \end{aligned}$$

By (69) we obtain

$$\begin{aligned} \|\hat{\Sigma}_{\hat{u}_r} - \Sigma_u\| &= \left\| \hat{\Sigma}_{\hat{H}} - \hat{\Phi}_r \hat{\Sigma}'_{\hat{H}, \hat{H}_r} - \Sigma_H + \Phi_r^* \Sigma'_{H, H_r} + \sum_{k=r+1}^{\infty} \Phi_k \mathbb{E} [H_{t-k}(\theta_0) H_t'(\theta_0)] \right\| \\ &= \left\| \hat{\Sigma}_{\hat{H}} - \Sigma_H - (\hat{\Phi}_r - \Phi_r^*) \hat{\Sigma}'_{\hat{H}, \hat{H}_r} - \Phi_r^* (\hat{\Sigma}'_{\hat{H}, \hat{H}_r} - \Sigma'_{H, H_r}) + \sum_{k=r+1}^{\infty} \Phi_k \mathbb{E} [H_{t-k}(\theta_0) H_t'(\theta_0)] \right\| \\ &\leq \|\hat{\Sigma}_{\hat{H}} - \Sigma_H\| + \|(\hat{\Phi}_r - \Phi_r^*) (\hat{\Sigma}'_{\hat{H}, \hat{H}_r} - \Sigma'_{H, H_r})\| + \|(\hat{\Phi}_r - \Phi_r^*) \Sigma'_{H, H_r}\| \\ &\quad + \|\Phi_r^* (\hat{\Sigma}'_{\hat{H}, \hat{H}_r} - \Sigma'_{H, H_r})\| + \left\| \sum_{k=r+1}^{\infty} \Phi_k \mathbb{E} [H_{t-k}(\theta_0) H_t'(\theta_0)] \right\|. \end{aligned} \quad (71)$$

From Lemma 9 and hypotheses of Theorem 3 (see (9)) we deduce that

$$\begin{aligned} \left\| \sum_{k=r+1}^{\infty} \Phi_k \mathbb{E} [H_{t-k}(\theta_0) H_t'(\theta_0)] \right\| &\leq \sum_{k=r+1}^{\infty} \|\Phi_k\| \left\| \mathbb{E} [H_{t-k}(\theta_0) H_t'(\theta_0)] \right\| \\ &\leq K \sum_{k=r+1}^{\infty} \rho^k \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0. \end{aligned}$$

Observe also that

$$\|\Phi_r^*\|^2 \leq \sum_{k \geq 1} \text{Tr}(\Phi_k \Phi_k') < \infty.$$

Therefore the convergence  $\hat{\Sigma}_{\hat{u}_r}$  to  $\Sigma_u$  will be a consequence of the four following properties:

- $\|\hat{\Sigma}_{\hat{H}} - \Sigma_H\| = o_{\mathbb{P}}(1)$ ,
- $\mathbb{P} - \lim_{n \rightarrow \infty} \|\hat{\Phi}_r - \Phi_r^*\| = 0$ ,
- $\mathbb{P} - \lim_{n \rightarrow \infty} \|\hat{\Sigma}'_{\hat{H}, \hat{H}_r} - \Sigma'_{H, H_r}\| = 0$  and
- $\|\Sigma'_{H, H_r}\| = O(1)$ .



The above properties will be proved thanks several lemmas that are stated and proved hereafter. This ends the proof of Theorem 3. For this, consider the following lemmas:

**Lemma 11.** *Under the assumptions of Theorem 3, we have*

$$\sup_{r \geq 1} \max \left\{ \|\Sigma_{H, \underline{H}_r}\|, \|\Sigma_{\underline{H}_r}\|, \|\Sigma_{\underline{H}_r}^{-1}\| \right\} < \infty.$$

*Proof.* See Lemma 1 in the supplementary material of [BMCF12]. □

**Lemma 12.** *Under the assumptions of Theorem 3 there exists a finite positive constant  $K$  such that, for  $1 \leq r_1, r_2 \leq r$  and  $1 \leq m_1, m_2 \leq p + q + 1$  we have*

$$\sup_{t \in \mathbb{Z}} \sum_{h=-\infty}^{\infty} |\text{Cov} \{H_{t-r_1, m_1}(\theta_0)H_{t-r_2, m_2}(\theta_0), H_{t-r_1-h, m_1}(\theta_0)H_{t-r_2-h, m_2}(\theta_0)\}| < K.$$

*Proof.* We denote in the sequel by  $\dot{\lambda}_{j,k}$  the coefficient  $\dot{\lambda}_{j,k}(\theta_0)$  defined in (31).

Using the fact that the process  $(H_t(\theta_0))_{t \in \mathbb{Z}}$  is centered and taking into consideration the strict stationarity of  $(\epsilon_t)_{t \in \mathbb{Z}}$  we obtain that for any  $t \in \mathbb{Z}$

$$\begin{aligned} & \sum_{h=-\infty}^{\infty} \left| \text{Cov}(H_{t-r_1, m_1}(\theta_0)H_{t-r_2, m_2}(\theta_0), H_{t-r_1-h, m_1}(\theta_0)H_{t-r_2-h, m_2}(\theta_0)) \right| \\ &= \sum_{h=-\infty}^{\infty} \left| \mathbb{E}[H_{t-r_1, m_1}(\theta_0)H_{t-r_2, m_2}(\theta_0)H_{t-r_1-h, m_1}(\theta_0)H_{t-r_2-h, m_2}(\theta_0)] \right. \\ & \quad \left. - \mathbb{E}[H_{t-r_1, m_1}(\theta_0)H_{t-r_2, m_2}(\theta_0)] \mathbb{E}[H_{t-r_1-h, m_1}(\theta_0)H_{t-r_2-h, m_2}(\theta_0)] \right| \\ &\leq \sum_{h=-\infty}^{\infty} \left| \text{cum}(H_{t-r_1, m_1}(\theta_0), H_{t-r_2, m_2}(\theta_0), H_{t-r_1-h, m_1}(\theta_0), H_{t-r_2-h, m_2}(\theta_0)) \right| \\ & \quad + \sum_{h=-\infty}^{\infty} |\mathbb{E}[H_{t-r_1, m_1}(\theta_0)H_{t-r_1-h, m_1}(\theta_0)]| |\mathbb{E}[H_{t-r_2, m_2}(\theta_0)H_{t-r_2-h, m_2}(\theta_0)]| \\ & \quad + \sum_{h=-\infty}^{\infty} |\mathbb{E}[H_{t-r_1, m_1}(\theta_0)H_{t-r_2-h, m_2}(\theta_0)]| |\mathbb{E}[H_{t-r_2, m_2}(\theta_0)H_{t-r_1-h, m_1}(\theta_0)]| \\ &\leq \sum_{h=-\infty}^{\infty} \sum_{i_1, j_1, k_1, \ell_1 \geq 1} \left| \dot{\lambda}_{i_1, m_1} \dot{\lambda}_{j_1, m_2} \dot{\lambda}_{k_1, m_1} \dot{\lambda}_{\ell_1, m_2} \right| |\text{cum}(\epsilon_0 \epsilon_{-i_1}, \epsilon_{r_1-r_2} \epsilon_{r_1-r_2-j_1}, \epsilon_{-h} \epsilon_{-h-k_1}, \epsilon_{r_1-r_2-h} \epsilon_{r_1-r_2-h-\ell_1})| \\ & \quad + T_{r1, m_1, r_2, m_2}^{(1)} + T_{r1, m_1, r_2, m_2}^{(2)}, \end{aligned}$$

where

$$T_{r1, m_1, r_2, m_2}^{(1)} = \sum_{h=-\infty}^{\infty} |\mathbb{E}[H_{t-r_1, m_1}(\theta_0)H_{t-r_1-h, m_1}(\theta_0)]| |\mathbb{E}[H_{t-r_2, m_2}(\theta_0)H_{t-r_2-h, m_2}(\theta_0)]|$$

and

$$T_{r1, m_1, r_2, m_2}^{(2)} = \sum_{h=-\infty}^{\infty} |\mathbb{E}[H_{t-r_1, m_1}(\theta_0)H_{t-r_2-h, m_2}(\theta_0)]| |\mathbb{E}[H_{t-r_2, m_2}(\theta_0)H_{t-r_1-h, m_1}(\theta_0)]|.$$

Thanks to Lemma 3 one may use the product theorem for the joint cumulants ([Bri81]) as in the proof of Lemma A.3. in [Sha11] in order to obtain that

$$\sum_{h=-\infty}^{\infty} \sum_{i_1, j_1, k_1, \ell_1 \geq 1} \left| \dot{\lambda}_{i_1, m_1} \dot{\lambda}_{j_1, m_2} \dot{\lambda}_{k_1, m_1} \dot{\lambda}_{\ell_1, m_2} \right| |\text{cum}(\epsilon_0 \epsilon_{-i_1}, \epsilon_{r_1-r_2} \epsilon_{r_1-r_2-j_1}, \epsilon_{-h} \epsilon_{-h-k_1}, \epsilon_{r_1-r_2-h} \epsilon_{r_1-r_2-h-\ell_1})| < \infty.$$

where we have used the absolute summability of the  $k$ -th ( $k = 2, \dots, 8$ ) cumulants assumed in (A4') with  $\tau = 8$ .

Observe now that

$$\begin{aligned} T_{r1, m_1, r_2, m_2}^{(1)} &= \sum_{h=-\infty}^{\infty} |\mathbb{E}[H_{t-r_1, m_1}(\theta_0) H_{t-r_1-h, m_1}(\theta_0)]| |\mathbb{E}[H_{t-r_2, m_2}(\theta_0) H_{t-r_2-h, m_2}(\theta_0)]| \\ &\leq \sup_{h \in \mathbb{Z}} |\mathbb{E}[H_{t-r_1, m_1}(\theta_0) H_{t-r_1-h, m_1}(\theta_0)]| \sum_{h=-\infty}^{\infty} |\mathbb{E}[H_{t-r_2, m_2}(\theta_0) H_{t-r_2-h, m_2}(\theta_0)]|. \end{aligned}$$

For any  $h \in \mathbb{Z}$ , from (30) we have

$$\begin{aligned} |\mathbb{E}[H_{t-r_1, m_1}(\theta_0) H_{t-r_1-h, m_1}(\theta_0)]| &\leq \sum_{i, j \geq 1} |\dot{\lambda}_{i, m_1}| |\dot{\lambda}_{j, m_1}| |\text{cum}(\epsilon_0, \epsilon_{-i}, \epsilon_{-h}, \epsilon_{-h-j})| \\ &\quad + \sum_{i, j \geq 1} |\dot{\lambda}_{i, m_1}| |\dot{\lambda}_{j, m_1}| \left\{ |\mathbb{E}[\epsilon_0 \epsilon_{-i}] \mathbb{E}[\epsilon_{-h} \epsilon_{-h-j}]| \right. \\ &\quad \left. + |\mathbb{E}[\epsilon_0 \epsilon_{-h}] \mathbb{E}[\epsilon_{-i} \epsilon_{-h-j}]| + |\mathbb{E}[\epsilon_0 \epsilon_{-h-j}] \mathbb{E}[\epsilon_{-i} \epsilon_{-h}]| \right\} \\ &\leq \sum_{i, j \geq 1} |\text{cum}(\epsilon_0, \epsilon_{-i}, \epsilon_{-h}, \epsilon_{-h-j})| + \sigma_\epsilon^4 \sum_{i \geq 1} |\dot{\lambda}_{i, m_1}|^2. \end{aligned}$$

Under Assumption (A4') with  $\tau = 4$  and in view of Lemma 3 we may write that

$$\sup_{h \in \mathbb{Z}} |\mathbb{E}[H_{t-r_1, m_1}(\theta_0) H_{t-r_1-h, m_1}(\theta_0)]| \leq \sup_{h \in \mathbb{Z}} \sum_{i, j \geq 1} |\text{cum}(\epsilon_0, \epsilon_{-i}, \epsilon_{-h}, \epsilon_{-h-j})| + \sigma_\epsilon^4 \sum_{i \geq 1} |\dot{\lambda}_{i, m_1}|^2 < \infty.$$

Similarly, we obtain

$$\begin{aligned} \sum_{h=-\infty}^{\infty} |\mathbb{E}[H_{t-r_2, m_2}(\theta_0) H_{t-r_2-h, m_2}(\theta_0)]| &\leq \sum_{h=-\infty}^{\infty} \sum_{i, j \geq 1} |\text{cum}(\epsilon_0, \epsilon_{-i}, \epsilon_{-h}, \epsilon_{-h-j})| + \sigma_\epsilon^4 \sum_{i \geq 1} |\dot{\lambda}_{i, m_1}|^2 \\ &< \infty. \end{aligned}$$

Consequently  $T_{r1, m_1, r_2, m_2}^{(1)} < \infty$  and the same approach yields that  $T_{r1, m_1, r_2, m_2}^{(2)} < \infty$  and the lemma is proved.  $\square$

Let  $\hat{\Sigma}_{\underline{H}_r}$ ,  $\hat{\Sigma}_{\underline{H}}$  and  $\hat{\Sigma}_{\underline{H}, \underline{H}_r}$  be the matrices obtained by replacing  $\hat{H}_t$  by  $H_t(\theta_0)$  in  $\hat{\Sigma}_{\hat{H}_r}$ ,  $\hat{\Sigma}_{\hat{H}}$  and  $\hat{\Sigma}_{\hat{H}, \hat{H}_r}$ .

**Lemma 13.** Under the assumptions of Theorem 3,  $\sqrt{r}\|\hat{\Sigma}_{\underline{H}_r} - \Sigma_{\underline{H}_r}\|$ ,  $\sqrt{r}\|\hat{\Sigma}_{H,\underline{H}_r} - \Sigma_{H,\underline{H}_r}\|$  and  $\sqrt{r}\|\hat{\Sigma}_H - \Sigma_H\|$  tend to zero in probability as  $n \rightarrow \infty$  when  $r = o(n^{1/3})$ .

*Proof.* For  $1 \leq m_1, m_2 \leq p + q + 1$  and  $1 \leq r_1, r_2 \leq r$ , the  $(\{(r_1 - 1)(p + q + 1) + m_1\}, \{(r_2 - 1)(p + q + 1) + m_2\})$ -th element of  $\hat{\Sigma}_{\underline{H}_r}$  is given by:

$$\frac{1}{n} \sum_{t=1}^n H_{t-r_1, m_1}(\theta_0) H_{t-r_2, m_2}(\theta_0).$$

For all  $\beta > 0$ , we use (65) and we obtain

$$\begin{aligned} \mathbb{P}\left(\sqrt{r}\|\hat{\Sigma}_{\underline{H}_r} - \Sigma_{\underline{H}_r}\| \geq \beta\right) &\leq \frac{r}{\beta^2} \mathbb{E}\left\|\hat{\Sigma}_{\underline{H}_r} - \Sigma_{\underline{H}_r}\right\|^2 \\ &\leq \frac{r}{\beta^2} \mathbb{E}\left\|\frac{1}{n} \sum_{t=1}^n \underline{H}_{r,t} \underline{H}'_{r,t} - \mathbb{E}[\underline{H}_{r,t} \underline{H}'_{r,t}]\right\|^2 \\ &\leq \frac{r}{\beta^2} \sum_{r_1=1}^r \sum_{r_2=1}^r \sum_{m_1=1}^{p+q+1} \sum_{m_2=1}^{p+q+1} \mathbb{E}\left(\frac{1}{n} \sum_{t=1}^n H_{t-r_1, m_1}(\theta_0) H_{t-r_2, m_2}(\theta_0) \right. \\ &\quad \left. - \mathbb{E}[H_{t-r_1, m_1}(\theta_0) H_{t-r_2, m_2}(\theta_0)]\right)^2. \end{aligned}$$

The stationarity of the process  $(H_{t-r_1, m_1}(\theta_0) H_{t-r_2, m_2}(\theta_0))_t$  and Lemma 12 imply

$$\begin{aligned} &\mathbb{P}\left(\sqrt{r}\|\hat{\Sigma}_{\underline{H}_r} - \Sigma_{\underline{H}_r}\| \geq \beta\right) \\ &\leq \frac{r}{\beta^2} \sum_{r_1=1}^r \sum_{r_2=1}^r \sum_{m_1=1}^{p+q+1} \sum_{m_2=1}^{p+q+1} \text{Var}\left(\frac{1}{n} \sum_{t=1}^n H_{t-r_1, m_1}(\theta_0) H_{t-r_2, m_2}(\theta_0)\right) \\ &\leq \frac{r}{(n\beta)^2} \sum_{r_1=1}^r \sum_{r_2=1}^r \sum_{m_1=1}^{p+q+1} \sum_{m_2=1}^{p+q+1} \sum_{t=1}^n \sum_{s=1}^n \text{Cov}(H_{t-r_1, m_1}(\theta_0) H_{t-r_2, m_2}(\theta_0), H_{s-r_1, m_1}(\theta_0) H_{s-r_2, m_2}(\theta_0)) \\ &\leq \frac{r}{(n\beta)^2} \sum_{r_1=1}^r \sum_{r_2=1}^r \sum_{m_1=1}^{p+q+1} \sum_{m_2=1}^{p+q+1} \sum_{h=1-n}^{n-1} (n - |h|) \text{Cov}(H_{t-r_1, m_1}(\theta_0) H_{t-r_2, m_2}(\theta_0), H_{t-h-r_1, m_1}(\theta_0) H_{t-h-r_2, m_2}(\theta_0)) \\ &\leq \frac{r}{n\beta^2} \sum_{r_1=1}^r \sum_{r_2=1}^r \sum_{m_1=1}^{p+q+1} \sum_{m_2=1}^{p+q+1} \sup_{t \in \mathbb{Z}} \sum_{h=-\infty}^{\infty} |\text{Cov}(H_{t-r_1, m_1}(\theta_0) H_{t-r_2, m_2}(\theta_0), H_{t-h-r_1, m_1}(\theta_0) H_{t-h-r_2, m_2}(\theta_0))| \\ &\leq \frac{C(p+q+1)^2 r^3}{n\beta^2}. \end{aligned}$$

Consequently we have

$$\begin{aligned} \mathbb{E}\left[r\|\hat{\Sigma}_H - \Sigma_H\|^2\right] &\leq \mathbb{E}\left[r\|\hat{\Sigma}_{H,\underline{H}_r} - \Sigma_{H,\underline{H}_r}\|^2\right] \\ &\leq \mathbb{E}\left[r\|\hat{\Sigma}_{\underline{H}_r} - \Sigma_{\underline{H}_r}\|^2\right] \\ &\leq \frac{C(p+q+1)^2 r^3}{n} \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

when  $r = o(n^{1/3})$ . The conclusion follows.  $\square$

We show in the following lemma that the previous lemma remains valid when we replace  $H_t(\theta_0)$  by  $\hat{H}_t$ .

**Lemma 14.** *Under the assumptions of Theorem 3,  $\sqrt{r}\|\hat{\Sigma}_{\hat{H}_r} - \Sigma_{H_r}\|$ ,  $\sqrt{r}\|\hat{\Sigma}_{\hat{H}_r, \hat{H}_r} - \Sigma_{H, H_r}\|$  and  $\sqrt{r}\|\hat{\Sigma}_{\hat{H}} - \Sigma_H\|$  tend to zero in probability as  $n \rightarrow \infty$  when  $r = o(n^{(1-2(d_2-d_1))/5})$ .*

*Proof.* As mentioned in the end of the proof of the previous lemma, we only have to deal with the term  $\sqrt{r}\|\hat{\Sigma}_{\hat{H}_r} - \Sigma_{H_r}\|$ .

We denote  $\hat{\Sigma}_{H_{r,n}}$  the matrix obtained by replacing  $\tilde{\epsilon}_t(\hat{\theta}_n)$  by  $\epsilon_t(\hat{\theta}_n)$  in  $\hat{\Sigma}_{\hat{H}_r}$ . We have

$$\sqrt{r}\|\hat{\Sigma}_{\hat{H}_r} - \Sigma_{H_r}\| \leq \sqrt{r}\|\hat{\Sigma}_{\hat{H}_r} - \hat{\Sigma}_{H_{r,n}}\| + \sqrt{r}\|\hat{\Sigma}_{H_{r,n}} - \hat{\Sigma}_{H_r}\| + \sqrt{r}\|\hat{\Sigma}_{H_r} - \Sigma_{H_r}\|.$$

By Lemma 13, the term  $\sqrt{r}\|\hat{\Sigma}_{H_r} - \Sigma_{H_r}\|$  converges in probability. The lemma will be proved as soon as we show that

$$\sqrt{r}\|\hat{\Sigma}_{\hat{H}_r} - \hat{\Sigma}_{H_{r,n}}\| = o_{\mathbb{P}}(1) \quad \text{and} \quad (72)$$

$$\sqrt{r}\|\hat{\Sigma}_{H_{r,n}} - \hat{\Sigma}_{H_r}\| = o_{\mathbb{P}}(1), \quad (73)$$

when  $r = o(n^{(1-2(d_2-d_1))/5})$ . This is done in two separate steps.

**Step 1: proof of (72).**

For all  $\beta > 0$ , we have

$$\begin{aligned} \mathbb{P}\left(\sqrt{r}\|\hat{\Sigma}_{\hat{H}_r} - \hat{\Sigma}_{H_{r,n}}\| \geq \beta\right) &\leq \frac{\sqrt{r}}{\beta} \mathbb{E}\|\hat{\Sigma}_{\hat{H}_r} - \hat{\Sigma}_{H_{r,n}}\| \\ &\leq \frac{\sqrt{r}}{\beta} \mathbb{E}\left\|\frac{1}{n} \sum_{t=1}^n \hat{H}_{r,t} \hat{H}_{r,t}' - \frac{1}{n} \sum_{t=1}^n H_{r,t}^{(n)} H_{r,t}^{(n)'}\right\| \\ &\leq \frac{K\sqrt{r}}{\beta} \sum_{r_1=1}^r \sum_{r_2=1}^r \sum_{m_1=1}^{p+q+1} \sum_{m_2=1}^{p+q+1} \mathbb{E}\left|\frac{1}{n} \sum_{t=1}^n \hat{H}_{t-r_1, m_1} \hat{H}_{t-r_2, m_2} - \frac{1}{n} \sum_{t=1}^n H_{t-r_1, m_1}^{(n)} H_{t-r_2, m_2}^{(n)}\right|, \end{aligned}$$

where

$$H_{t,m}^{(n)} = 2\epsilon_t(\hat{\theta}_n) \frac{\partial}{\partial \theta_m} \epsilon_t(\hat{\theta}_n) \quad \text{and} \quad H_{r,t}^{(n)} = \left(H_{t-1}^{(n)'}, \dots, H_{t-r}^{(n)'}\right)'.$$

It follows that

$$\begin{aligned} \mathbb{P}\left(\sqrt{r}\|\hat{\Sigma}_{\hat{H}_r} - \hat{\Sigma}_{H_{r,n}}\| \geq \beta\right) &\leq \frac{4K\sqrt{r}}{n\beta} \sum_{r_1=1}^r \sum_{r_2=1}^r \sum_{m_1=1}^{p+q+1} \sum_{m_2=1}^{p+q+1} \\ &\quad \mathbb{E}\left|\sum_{t=1}^n \tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_1}} \tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_2}} \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \right. \\ &\quad \left. - \epsilon_{t-r_1}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \epsilon_{t-r_2}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n)\right|. \end{aligned} \quad (74)$$

Observe now that

$$\begin{aligned}
& \tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_1}} \tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_2}} \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) - \epsilon_{t-r_1}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \epsilon_{t-r_2}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n) \\
&= \left( \tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) - \epsilon_{t-r_1}(\hat{\theta}_n) \right) \frac{\partial}{\partial \theta_{m_1}} \tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_2}} \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \\
&\quad + \epsilon_{t-r_1}(\hat{\theta}_n) \left( \frac{\partial}{\partial \theta_{m_1}} \tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) - \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \right) \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_2}} \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \\
&\quad + \epsilon_{t-r_1}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \left( \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) - \epsilon_{t-r_2}(\hat{\theta}_n) \right) \frac{\partial}{\partial \theta_{m_2}} \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \\
&\quad + \epsilon_{t-r_1}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \epsilon_{t-r_2}(\hat{\theta}_n) \left( \frac{\partial}{\partial \theta_{m_2}} \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) - \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n) \right).
\end{aligned}$$

We replace the above identity in (74) and we obtain by Hölder's inequality that

$$\mathbb{P} \left( \sqrt{r} \left\| \hat{\Sigma}_{\underline{H}_r} - \hat{\Sigma}_{\underline{H}_{r,n}} \right\| \geq \beta \right) \leq \frac{4K\sqrt{r}}{n\beta} \sum_{r_1=1}^r \sum_{r_2=1}^r \sum_{m_1=1}^{p+q+1} \sum_{m_2=1}^{p+q+1} (T_{n,1} + T_{n,2} + T_{n,3} + T_{n,4}) \quad (75)$$

where

$$\begin{aligned}
T_{n,1} &= \sum_{t=1}^n \left\| \tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) - \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^2} \left\| \frac{\partial}{\partial \theta_{m_1}} \tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \frac{\partial}{\partial \theta_{m_2}} \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^6}, \\
T_{n,2} &= \sum_{t=1}^n \left\| \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \frac{\partial}{\partial \theta_{m_1}} \tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) - \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^2} \left\| \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \frac{\partial}{\partial \theta_{m_2}} \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^6}, \\
T_{n,3} &= \sum_{t=1}^n \left\| \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) - \epsilon_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^2} \left\| \frac{\partial}{\partial \theta_{m_2}} \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^6}, \\
T_{n,4} &= \sum_{t=1}^n \left\| \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \epsilon_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \frac{\partial}{\partial \theta_{m_2}} \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) - \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^2}.
\end{aligned}$$

For all  $\theta \in \Theta_\delta$  and  $t \in \mathbb{Z}$ , in view of (30) and Lemma 1, we have

$$\begin{aligned}
\left\| \tilde{\epsilon}_t(\hat{\theta}_n) - \epsilon_t(\hat{\theta}_n) \right\|_{\mathbb{L}^2} &= \left( \mathbb{E} \left[ \left\{ \sum_{j \geq 0} \left( \lambda_j^t(\hat{\theta}_n) - \lambda_j(\hat{\theta}_n) \right) \epsilon_{t-j} \right\}^2 \right] \right)^{1/2} \\
&\leq \sup_{\theta \in \Theta_\delta} \left( \mathbb{E} \left[ \left\{ \sum_{j \geq 0} \left( \lambda_j^t(\theta) - \lambda_j(\theta) \right) \epsilon_{t-j} \right\}^2 \right] \right)^{1/2} \\
&\leq \sigma_\epsilon \sup_{\theta \in \Theta_\delta} \left\| \lambda(\theta) - \lambda^t(\theta) \right\|_{\ell^2} \\
&\leq K \frac{1}{t^{1/2-(d_2-d_1)}}.
\end{aligned}$$

It is not difficult to prove that  $\tilde{\epsilon}_t(\theta)$  and  $\partial \tilde{\epsilon}_t(\theta)/\partial \theta$  belong to  $\mathbb{L}^6$ . The fact that  $\epsilon_t(\theta)$  and  $\partial \epsilon_t(\theta)/\partial \theta$  have moment of order 6 can be proved using the same method than in Lemma 12 using the absolute

summability of the  $k$ -th ( $k = 2, \dots, 8$ ) cumulants assumed in **(A4')** with  $\tau = 8$ . We deduce that

$$\begin{aligned} T_{n,1} &\leq K \sum_{t=1}^n \left\| \tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) - \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^2} \leq K \sum_{t=1-r}^0 \left\| \epsilon_t(\hat{\theta}_n) \right\|_{\mathbb{L}^2} + K \sum_{t=1}^n \left\| \tilde{\epsilon}_t(\hat{\theta}_n) - \epsilon_t(\hat{\theta}_n) \right\|_{\mathbb{L}^2} \\ &\leq K \left( r + \sum_{t=1}^n \frac{1}{t^{1/2-(d_2-d_1)}} \right). \end{aligned}$$

Then we obtain

$$T_{n,1} \leq K \left( r + n^{1/2+(d_2-d_1)} \right). \quad (76)$$

The same calculations hold for the terms  $T_{n,2}$ ,  $T_{n,3}$  and  $T_{n,4}$ . Thus

$$T_{n,1} + T_{n,2} + T_{n,3} + T_{n,4} \leq K \left( r + n^{1/2+(d_2-d_1)} \right) \quad (77)$$

and reporting this estimation in (75) implies that

$$\begin{aligned} \mathbb{P} \left( \sqrt{r} \left\| \hat{\Sigma}_{\underline{H}_r} - \hat{\Sigma}_{\underline{H}_{r,n}} \right\| \geq \beta \right) &\leq \frac{K r^{5/2} (p+q+1)^2}{n\beta} \left( r + n^{1/2+(d_2-d_1)} \right) \\ &\leq K \left( \frac{r^{7/2}}{n} + \frac{r^{5/2}}{n^{1/2-(d_2-d_1)}} \right). \end{aligned}$$

Since  $2/7 > (1 - 2(d_2 - d_1))/5$ , the sequence  $\sqrt{r} \left\| \hat{\Sigma}_{\underline{H}_r} - \hat{\Sigma}_{\underline{H}_{r,n}} \right\|$  converges in probability to 0 as  $n \rightarrow \infty$  when  $r = r(n) = o(n^{(1-2(d_2-d_1))/5})$ .

### Step 2: proof of (73).

First we follow the same approach than in the previous step. We have

$$\begin{aligned} \left\| \hat{\Sigma}_{\underline{H}_{r,n}} - \hat{\Sigma}_{\underline{H}_r} \right\|^2 &= \left\| \frac{1}{n} \sum_{t=1}^n \underline{H}_{r,t}^{(n)} \underline{H}_{r,t}^{(n)'} - \frac{1}{n} \sum_{t=1}^n \underline{H}_{r,t} \underline{H}_{r,t}' \right\|^2 \\ &\leq \sum_{r_1=1}^r \sum_{r_2=1}^r \sum_{m_1=1}^{p+q+1} \sum_{m_2=1}^{p+q+1} \left( \frac{1}{n} \sum_{t=1}^n H_{t-r_1, m_1}^{(n)} H_{t-r_2, m_2}^{(n)} - \frac{1}{n} \sum_{t=1}^n H_{t-r_1, m_1} H_{t-r_2, m_2} \right)^2 \\ &\leq 16 \sum_{r_1=1}^r \sum_{r_2=1}^r \sum_{m_1=1}^{p+q+1} \sum_{m_2=1}^{p+q+1} \left( \frac{1}{n} \sum_{t=1}^n \epsilon_{t-r_1}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \epsilon_{t-r_2}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n) \right. \\ &\quad \left. - \epsilon_{t-r_1}(\theta_0) \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\theta_0) \epsilon_{t-r_2}(\theta_0) \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\theta_0) \right)^2. \end{aligned}$$

Since

$$\begin{aligned}
& \epsilon_{t-r_1}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \epsilon_{t-r_2}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n) - \epsilon_{t-r_1}(\theta_0) \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\theta_0) \epsilon_{t-r_2}(\theta_0) \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\theta_0) \\
&= \left( \epsilon_{t-r_1}(\hat{\theta}_n) - \epsilon_{t-r_1}(\theta_0) \right) \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \epsilon_{t-r_2}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n) \\
&\quad + \epsilon_{t-r_1}(\theta_0) \left( \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) - \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\theta_0) \right) \epsilon_{t-r_2}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n) \\
&\quad + \epsilon_{t-r_1}(\theta_0) \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\theta_0) \left( \epsilon_{t-r_2}(\hat{\theta}_n) - \epsilon_{t-r_2}(\theta_0) \right) \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n) \\
&\quad + \epsilon_{t-r_1}(\theta_0) \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\theta_0) \epsilon_{t-r_2}(\theta_0) \left( \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n) - \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\theta_0) \right),
\end{aligned}$$

one has

$$\left\| \hat{\Sigma}_{H_{r,n}} - \hat{\Sigma}_{H_r} \right\|^2 \leq 16 \sum_{r_1=1}^r \sum_{r_2=1}^r \sum_{m_1=1}^{p+q+1} \sum_{m_2=1}^{p+q+1} (U_{n,1} + U_{n,2} + U_{n,3} + U_{n,4})^2 \quad (78)$$

where

$$\begin{aligned}
U_{n,1} &= \frac{1}{n} \sum_{t=1}^n \left| \epsilon_{t-r_1}(\hat{\theta}_n) - \epsilon_{t-r_1}(\theta_0) \right| \left| \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \right| \left| \epsilon_{t-r_2}(\hat{\theta}_n) \right| \left| \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n) \right|, \\
U_{n,2} &= \frac{1}{n} \sum_{t=1}^n \left| \epsilon_{t-r_1}(\theta_0) \right| \left| \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) - \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\theta_0) \right| \left| \epsilon_{t-r_2}(\hat{\theta}_n) \right| \left| \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n) \right|, \\
U_{n,3} &= \frac{1}{n} \sum_{t=1}^n \left| \epsilon_{t-r_1}(\theta_0) \right| \left| \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\theta_0) \right| \left| \epsilon_{t-r_2}(\hat{\theta}_n) - \epsilon_{t-r_2}(\theta_0) \right| \left| \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n) \right| \\
U_{n,4} &= \frac{1}{n} \sum_{t=1}^n \left| \epsilon_{t-r_1}(\theta_0) \right| \left| \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\theta_0) \right| \left| \epsilon_{t-r_2}(\theta_0) \right| \left| \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n) - \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\theta_0) \right|.
\end{aligned}$$

Taylor expansions around  $\theta_0$  yield that there exists  $\underline{\theta}$  and  $\bar{\theta}$  between  $\hat{\theta}_n$  and  $\theta_0$  such that

$$\begin{aligned}
\left| \epsilon_t(\hat{\theta}_n) - \epsilon_t(\theta_0) \right| &\leq w_t \left\| \hat{\theta}_n - \theta_0 \right\| \quad \text{and} \\
\left| \frac{\partial}{\partial \theta_m} \epsilon_t(\hat{\theta}_n) - \frac{\partial}{\partial \theta_m} \epsilon_t(\theta_0) \right| &\leq q_t \left\| \hat{\theta}_n - \theta_0 \right\|
\end{aligned}$$

with  $w_t = \left\| \partial \epsilon_t(\underline{\theta}) / \partial \theta' \right\|$  and  $q_t = \left\| \partial^2 \epsilon_t(\bar{\theta}) / \partial \theta' \partial \theta_m \right\|$ . Using the fact that

$$\mathbb{E} \left| w_{t-r_1} \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \epsilon_{t-r_2}(\hat{\theta}_n) \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n) \right| < \infty$$

and that  $(\sqrt{n}(\hat{\theta}_n - \theta_0))_n$  is a tight sequence (which implies that  $\|\hat{\theta}_n - \theta_0\| = O_{\mathbb{P}}(1/\sqrt{n})$ ), we deduce that

$$U_{n,1} = O_{\mathbb{P}} \left( \frac{1}{\sqrt{n}} \right).$$

The same arguments are valid for  $U_{n,2}$ ,  $U_{n,3}$  and  $U_{n,4}$ . Consequently  $U_{n,1} + U_{n,2} + U_{n,3} + U_{n,4} = O_{\mathbb{P}}(1/\sqrt{n})$  and (78) yields

$$\left\| \hat{\Sigma}_{\underline{H}_{r,n}} - \hat{\Sigma}_{\underline{H}_r} \right\|^2 = O_{\mathbb{P}}\left(\frac{r^2}{n}\right).$$

When  $r = o(n^{1/3})$  we finally obtain  $\sqrt{r} \|\hat{\Sigma}_{\underline{H}_{r,n}} - \hat{\Sigma}_{\underline{H}_r}\| = o_{\mathbb{P}}(1)$ .

□

**Lemma 15.** *Under the assumptions of Theorem 3, we have*

$$\sqrt{r} \|\underline{\Phi}_r^* - \underline{\Phi}_r\| = o_{\mathbb{P}}(1) \quad \text{as } r \rightarrow \infty.$$

*Proof.* Recall that by (8) and (68) we have

$$H_t(\theta_0) = \underline{\Phi}_r \underline{H}_{r,t} + u_{r,t} = \underline{\Phi}_r^* \underline{H}_{r,t} + \sum_{k=r+1}^{\infty} \phi_k H_{t-k}(\theta_0) + u_t := \underline{\Phi}_r^* \underline{H}_{r,t} + u_{r,t}^*.$$

By the orthogonality conditions in (8) and (68), one has

$$\begin{aligned} \Sigma_{u_r^*, \underline{H}_r} &:= \mathbb{E} \left[ u_{r,t}^* \underline{H}_{r,t}' \right] = \mathbb{E} \left[ (H_t(\theta_0) - \underline{\Phi}_r^* \underline{H}_{r,t}) \underline{H}_{r,t}' \right] = \mathbb{E} \left[ (\underline{\Phi}_r \underline{H}_{r,t} + u_{r,t} - \underline{\Phi}_r^* \underline{H}_{r,t}) \underline{H}_{r,t}' \right] \\ &= (\underline{\Phi}_r - \underline{\Phi}_r^*) \Sigma_{\underline{H}_r}, \end{aligned}$$

and consequently

$$\underline{\Phi}_r^* - \underline{\Phi}_r = -\Sigma_{u_r^*, \underline{H}_r} \Sigma_{\underline{H}_r}^{-1}. \quad (79)$$

Using Lemma 11 and Lemma 12, (79) implies that

$$\begin{aligned} \mathbb{P}(\sqrt{r} \|\underline{\Phi}_r^* - \underline{\Phi}_r\| \geq \beta) &\leq \frac{\sqrt{r}}{\beta} \|\Sigma_{u_r^*, \underline{H}_r}\| \|\Sigma_{\underline{H}_r}^{-1}\| \\ &\leq \frac{K\sqrt{r}}{\beta} \left\| \mathbb{E} \left[ \left( \sum_{k \geq r+1} \phi_k H_{t-k}(\theta_0) + u_t \right) \underline{H}_{r,t}' \right] \right\| \\ &\leq \frac{K\sqrt{r}}{\beta} \sum_{k \geq r+1} \|\phi_k\| \left\| \mathbb{E} \left[ H_{t-k}(\theta_0) \underline{H}_{r,t}' \right] \right\| \\ &\leq \frac{K\sqrt{r}}{\beta} \sum_{\ell \geq 1} \|\phi_{\ell+r}\| \left\| \mathbb{E} \left[ H_{t-\ell-r}(\theta_0) (H'_{t-1}(\theta_0), \dots, H'_{t-r}(\theta_0)) \right] \right\| \\ &\leq \frac{K\sqrt{r}}{\beta} \sum_{\ell \geq 1} \|\phi_{\ell+r}\| \left( \sum_{j=1}^{p+q+1} \sum_{k=1}^{p+q+1} \sum_{r_1=1}^r \left| \mathbb{E} [H_{t-r-\ell,j}(\theta_0) H_{t-r_1,k}(\theta_0)] \right|^2 \right)^{1/2} \\ &\leq \frac{K\sqrt{r}}{\beta} \sum_{\ell \geq 1} \|\phi_{\ell+r}\| \left( \sum_{j=1}^{p+q+1} \sum_{k=1}^{p+q+1} \sum_{r_1=1}^r \mathbb{E} [H_{t-r-\ell,j}^2(\theta_0)] \mathbb{E} [H_{t-r_1,k}^2(\theta_0)] \right)^{1/2} \\ &\leq \frac{K(p+q+1)r}{\beta} \sum_{\ell \geq 1} \|\phi_{\ell+r}\|. \end{aligned}$$

By (9),  $r \sum_{\ell \geq 1} \|\phi_{\ell+r}\| = o(1)$  as  $r \rightarrow \infty$ . The proof of the lemma follows.

□



**Lemma 16.** *Under the assumptions of Theorem 3, we have*

$$\sqrt{r} \left\| \hat{\Sigma}_{\underline{H}_r}^{-1} - \Sigma_{\underline{H}_r}^{-1} \right\| = o_{\mathbb{P}}(1)$$

as  $n \rightarrow \infty$  when  $r = o(n^{(1-2(d_2-d_1))/5})$  and  $r \rightarrow \infty$ .

*Proof.* We have

$$\left\| \hat{\Sigma}_{\underline{H}_r}^{-1} - \Sigma_{\underline{H}_r}^{-1} \right\| \leq \left( \left\| \hat{\Sigma}_{\underline{H}_r}^{-1} - \Sigma_{\underline{H}_r}^{-1} \right\| + \left\| \Sigma_{\underline{H}_r}^{-1} \right\| \right) \left\| \Sigma_{\underline{H}_r} - \hat{\Sigma}_{\underline{H}_r} \right\| \left\| \Sigma_{\underline{H}_r}^{-1} \right\|,$$

and by induction we obtain

$$\left\| \hat{\Sigma}_{\underline{H}_r}^{-1} - \Sigma_{\underline{H}_r}^{-1} \right\| \leq \left\| \Sigma_{\underline{H}_r}^{-1} \right\| \sum_{k=1}^{\infty} \left\| \Sigma_{\underline{H}_r} - \hat{\Sigma}_{\underline{H}_r} \right\|^k \left\| \Sigma_{\underline{H}_r}^{-1} \right\|^k.$$

We have

$$\begin{aligned} & \mathbb{P} \left( \sqrt{r} \left\| \hat{\Sigma}_{\underline{H}_r}^{-1} - \Sigma_{\underline{H}_r}^{-1} \right\| > \beta \right) \\ & \leq \mathbb{P} \left( \sqrt{r} \left\| \Sigma_{\underline{H}_r}^{-1} \right\| \sum_{k=1}^{\infty} \left\| \Sigma_{\underline{H}_r} - \hat{\Sigma}_{\underline{H}_r} \right\|^k \left\| \Sigma_{\underline{H}_r}^{-1} \right\|^k > \beta \right) \\ & \leq \mathbb{P} \left( \sqrt{r} \left\| \Sigma_{\underline{H}_r}^{-1} \right\| \sum_{k=1}^{\infty} \left\| \Sigma_{\underline{H}_r} - \hat{\Sigma}_{\underline{H}_r} \right\|^k \left\| \Sigma_{\underline{H}_r}^{-1} \right\|^k > \beta \text{ and } \left\| \Sigma_{\underline{H}_r} - \hat{\Sigma}_{\underline{H}_r} \right\| \left\| \Sigma_{\underline{H}_r}^{-1} \right\| < 1 \right) \\ & + \mathbb{P} \left( \sqrt{r} \left\| \Sigma_{\underline{H}_r}^{-1} \right\| \sum_{k=1}^{\infty} \left\| \Sigma_{\underline{H}_r} - \hat{\Sigma}_{\underline{H}_r} \right\|^k \left\| \Sigma_{\underline{H}_r}^{-1} \right\|^k > \beta \text{ and } \left\| \Sigma_{\underline{H}_r} - \hat{\Sigma}_{\underline{H}_r} \right\| \left\| \Sigma_{\underline{H}_r}^{-1} \right\| \geq 1 \right) \\ & \leq \mathbb{P} \left( \sqrt{r} \frac{\left\| \Sigma_{\underline{H}_r}^{-1} \right\|^2 \left\| \Sigma_{\underline{H}_r} - \hat{\Sigma}_{\underline{H}_r} \right\|}{1 - \left\| \Sigma_{\underline{H}_r} - \hat{\Sigma}_{\underline{H}_r} \right\| \left\| \Sigma_{\underline{H}_r}^{-1} \right\|} > \beta \right) + \mathbb{P} \left( \sqrt{r} \left\| \Sigma_{\underline{H}_r} - \hat{\Sigma}_{\underline{H}_r} \right\| \left\| \Sigma_{\underline{H}_r}^{-1} \right\| \geq 1 \right) \\ & \leq \mathbb{P} \left( \sqrt{r} \left\| \Sigma_{\underline{H}_r} - \hat{\Sigma}_{\underline{H}_r} \right\| > \frac{\beta}{\left\| \Sigma_{\underline{H}_r}^{-1} \right\|^2 + \beta r^{-1/2} \left\| \Sigma_{\underline{H}_r}^{-1} \right\|} \right) \\ & \quad + \mathbb{P} \left( \sqrt{r} \left\| \Sigma_{\underline{H}_r} - \hat{\Sigma}_{\underline{H}_r} \right\| \geq \left\| \Sigma_{\underline{H}_r}^{-1} \right\|^{-1} \right). \end{aligned}$$

Lemmas 11 and 14 imply the result.  $\square$

**Lemma 17.** *Under the assumptions of Theorem 3, we have*

$$\sqrt{r} \left\| \hat{\underline{\Phi}}_r - \underline{\Phi}_r \right\| = o_{\mathbb{P}}(1) \quad \text{as } r \rightarrow \infty \text{ and } r = o(n^{(1-2(d_2-d_1))/5}).$$

*Proof.* Lemmas 11 and 16 yield

$$\left\| \hat{\Sigma}_{\underline{H}_r}^{-1} \right\| \leq \left\| \hat{\Sigma}_{\underline{H}_r}^{-1} - \Sigma_{\underline{H}_r}^{-1} \right\| + \left\| \Sigma_{\underline{H}_r}^{-1} \right\| = O_{\mathbb{P}}(1). \quad (80)$$

By (68), we have

$$0 = \mathbb{E} \left[ u_{r,t} \underline{H}'_{r,t} \right] = \mathbb{E} \left[ (H_t(\theta_0) - \underline{\Phi}_r \underline{H}_{r,t}) \underline{H}'_{r,t} \right] = \Sigma_{H, \underline{H}_r} - \underline{\Phi}_r \Sigma_{\underline{H}_r},$$

and so we have  $\underline{\phi}_r = \Sigma_{H, \underline{H}_r} \Sigma_{\underline{H}_r}^{-1}$ . Lemmas 11, 14, 16 and (80) imply

$$\begin{aligned} \sqrt{r} \left\| \hat{\underline{\phi}}_r - \underline{\phi}_r \right\| &= \sqrt{r} \left\| \hat{\Sigma}_{\hat{H}, \hat{\underline{H}}_r} \hat{\Sigma}_{\hat{\underline{H}}_r}^{-1} - \Sigma_{H, \underline{H}_r} \Sigma_{\underline{H}_r}^{-1} \right\| \\ &= \sqrt{r} \left\| \left( \hat{\Sigma}_{\hat{H}, \hat{\underline{H}}_r} - \Sigma_{H, \underline{H}_r} \right) \hat{\Sigma}_{\hat{\underline{H}}_r}^{-1} + \Sigma_{H, \underline{H}_r} \left( \hat{\Sigma}_{\hat{\underline{H}}_r}^{-1} - \Sigma_{\underline{H}_r}^{-1} \right) \right\| \\ &= o_{\mathbb{P}}(1), \end{aligned}$$

and the lemma is proved.  $\square$

### Proof of Theorem 3

Since by Lemma 14 we have  $\|\hat{\Sigma}_{\hat{H}} - \Sigma_H\| = o_{\mathbb{P}}(r^{-1/2}) = o_{\mathbb{P}}(1)$  and  $\|\hat{\Sigma}_{\hat{H}, \hat{\underline{H}}_r} - \Sigma_{H, \underline{H}_r}\| = o_{\mathbb{P}}(r^{-1/2}) = o_{\mathbb{P}}(1)$ , and by Lemma 15  $\|\hat{\underline{\phi}}_r - \underline{\phi}_r^*\| = o_{\mathbb{P}}(r^{-1/2}) = o_{\mathbb{P}}(1)$ , Theorem 3 is then proved.

### 6.5. Invertibility of the normalization matrix $P_{p+q+1, n}$

The following proofs are quite technical and are adaptations of the arguments used in [BMS18].

To prove Proposition 4, we need to introduce the following notation.

We denote  $S_t$  the vector of  $\mathbb{R}^{p+q+1}$  defined by:

$$S_t = \sum_{j=1}^t U_j = \sum_{j=1}^t -2J^{-1}H_j = -2J^{-1} \sum_{j=1}^t \epsilon_j \frac{\partial}{\partial \theta} \epsilon_j(\theta_0),$$

and  $S_t(i)$  is its  $i$ -th component. We have

$$S_{t-1}(i) = S_t(i) - U_t(i). \quad (81)$$

If the matrix  $P_{p+q+1, n}$  is not invertible, there exists some real constants  $d_1, \dots, d_{p+q+1}$  not all equal to zero, such that  $\mathbf{d}' P_{p+q+1, n} \mathbf{d} = 0$ , where  $\mathbf{d} = (d_1, \dots, d_{p+q+1})'$ . Thus we may write that  $\sum_{i=1}^{p+q+1} \sum_{j=1}^{p+q+1} d_j P_{p+q+1, n}(j, i) d_i = 0$  or equivalently

$$\frac{1}{n^2} \sum_{t=1}^n \sum_{i=1}^{p+q+1} \sum_{j=1}^{p+q+1} d_j \left( \sum_{k=1}^t (U_k(j) - \bar{U}_n(j)) \right) \left( \sum_{k=1}^t (U_k(i) - \bar{U}_n(i)) \right) d_i = 0.$$

Then

$$\sum_{t=1}^n \left( \sum_{i=1}^{p+q+1} d_i \left( \sum_{k=1}^t (U_k(i) - \bar{U}_n(i)) \right) \right)^2 = 0,$$

which implies that for all  $t \geq 1$

$$\sum_{i=1}^{p+q+1} d_i \left( \sum_{k=1}^t (U_k(i) - \bar{U}_n(i)) \right) = \sum_{i=1}^{p+q+1} d_i \left( S_t(i) - \frac{t}{n} S_n(i) \right) = 0.$$

So we have

$$\frac{1}{t} \sum_{i=1}^{p+q+1} d_i S_t(i) = \sum_{i=1}^{p+q+1} d_i \left( \frac{1}{n} S_n(i) \right). \quad (82)$$

We apply the ergodic theorem and we use the orthogonality of  $\epsilon_t$  and  $(\partial/\partial\theta)\epsilon_t(\theta_0)$  in order to obtain that

$$\sum_{i=1}^{p+q+1} d_i \left( \frac{1}{n} \sum_{k=1}^n U_k(i) \right) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \sum_{i=1}^{p+q+1} d_i \mathbb{E}[U_k(i)] = -2 \sum_{i,j=1}^{p+q+1} d_i J^{-1}(i,j) \mathbb{E} \left[ \epsilon_k \frac{\partial \epsilon_k}{\partial \theta_j} \right] = 0.$$

Reporting this convergence in (82) implies that  $\sum_{i=1}^{p+q+1} d_i S_t(i) = 0$  a.s. for all  $t \geq 1$ . By (81), we deduce that

$$\sum_{i=1}^{p+q+1} d_i U_t(i) = -2 \sum_{i=1}^{p+q+1} d_i \sum_{j=1}^{p+q+1} J^{-1}(i,j) \left( \epsilon_t \frac{\partial \epsilon_t}{\partial \theta_j} \right) = 0, \quad \text{a.s.}$$

Thanks to Assumption (A5),  $(\epsilon_t)_{t \in \mathbb{Z}}$  has a positive density in some neighborhood of zero and then  $\epsilon_t \neq 0$  almost-surely. So we would have  $\mathbf{d}' J^{-1} \frac{\partial \epsilon_t}{\partial \theta} = 0$  a.s. Now we can follow the same arguments that we developed in the proof of the invertibility of  $J$  (see Proof of Lemma 6 and more precisely (53)) and this leads us to a contradiction. We deduce that the matrix  $P_{p+q+1,n}$  is non singular.

### 6.6. Proof of Theorem 5

The arguments follows the one [BMS18] in a simpler context.

Recall that the Skorohod space  $\mathbb{D}^\ell[0,1]$  is the set of  $\mathbb{R}^\ell$ -valued functions on  $[0,1]$  which are right-continuous and has left limits everywhere. It is endowed with the Skorohod topology and the weak convergence on  $\mathbb{D}^\ell[0,1]$  is mentioned by  $\xrightarrow{\mathbb{D}^\ell}$ . The integer part of  $x$  will be denoted by  $\lfloor x \rfloor$ .

The goal at first is to show that there exists a lower triangular matrix  $T$  with nonnegative diagonal entries such that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} U_t \xrightarrow[n \rightarrow \infty]{\mathbb{D}^{p+q+1}} (T T')^{1/2} B_{p+q+1}(r), \quad (83)$$

where  $(B_{p+q+1}(r))_{r \geq 0}$  is a  $(p+q+1)$ -dimensional standard Brownian motion. Using (30),  $U_t$  can be rewritten as

$$U_t = \left( -2 \left\{ \sum_{i=1}^{\infty} \dot{\lambda}_{i,1}(\theta_0) \epsilon_t \epsilon_{t-i}, \dots, \sum_{i=1}^{\infty} \dot{\lambda}_{i,p+q+1}(\theta_0) \epsilon_t \epsilon_{t-i} \right\} J^{-1'} \right)'.$$

The non-correlation between  $\epsilon_t$ 's implies that the process  $(U_t)_{t \in \mathbb{Z}}$  is centered. In order to apply the functional central limit theorem for strongly mixing process, we need to identify the asymptotic covariance matrix in the classical central limit theorem for the sequence  $(U_t)_{t \in \mathbb{Z}}$ . It is proved in Theorem 2 that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n U_t \xrightarrow[n \rightarrow \infty]{\text{in law}} \mathcal{N}(0, \Omega =: 2\pi f_U(0)),$$

where  $f_U(0)$  is the spectral density of the stationary process  $(U_t)_{t \in \mathbb{Z}}$  evaluated at frequency 0. The existence of the matrix  $\Omega$  has already been discussed (see the proofs of lemmas 6 and 9).

Since the matrix  $\Omega$  is positive definite, it can be factorized as  $\Omega = T T'$  where the  $(p+q+1) \times (p+q+1)$  lower triangular matrix  $T$  has nonnegative diagonal entries. Therefore, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n (T T')^{-1/2} F_t \xrightarrow[n \rightarrow \infty]{\text{in law}} \mathcal{N}(0, I_{p+q+1}),$$

where  $(TT')^{-1/2}$  is the Moore-Penrose inverse (see [MN99], p. 36) of  $(TT')^{1/2}$  and  $I_{p+q+1}$  is the identity matrix of order  $p+q+1$ .

As in the proof of the asymptotic normality of  $(\sqrt{n}(\hat{\theta}_n - \theta_0))_n$ , the distribution of  $n^{-1/2} \sum_{t=1}^n U_t$  when  $n$  tends to infinity is obtained by introducing the random vector  $U_t^k$  defined for any strictly positive integer  $k$  by

$$U_t^k = \left( -2 \left\{ \sum_{i=1}^k \dot{\lambda}_{i,1}(\theta_0) \epsilon_t \epsilon_{t-i}, \dots, \sum_{i=1}^k \dot{\lambda}_{i,p+q+1}(\theta_0) \epsilon_t \epsilon_{t-i} \right\} J^{-1'} \right)'.$$

Since  $U_t^k$  depends on a finite number of values of the noise-process  $(\epsilon_t)_{t \in \mathbb{Z}}$ , it also satisfies a mixing property (see Theorem 14.1 in [Dav94], p. 210). The central limit theorem for strongly mixing process of [Her84] shows that its asymptotic distribution is normal with zero mean and variance matrix  $\Omega_k$  that converges when  $k$  tends to infinity to  $\Omega$  (see the proof of Lemma 10):

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n U_t^k \xrightarrow[n \rightarrow \infty]{\text{in law}} \mathcal{N}(0, \Omega_k).$$

The above arguments also apply to matrix  $\Omega_k$  with some matrix  $T_k$  which is defined analogously as  $T$ . Consequently, we obtain

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n (T_k T_k')^{-1/2} U_t^k \xrightarrow[n \rightarrow \infty]{\text{in law}} \mathcal{N}(0, I_{p+q+1}).$$

Now we are able to apply the functional central limit theorem for strongly mixing process of [Her84] and we obtain that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} (T_k T_k')^{-1/2} U_t^k \xrightarrow[n \rightarrow \infty]{\mathbb{D}^{p+q+1}} B_{p+q+1}(r).$$

Since

$$(TT')^{-1/2} U_t^k = \left( (TT')^{-1/2} - (T_k T_k')^{-1/2} \right) U_t^k + (T_k T_k')^{-1/2} U_t^k,$$

we may use the same approach as in the proof of Lemma 10 in order to prove that  $n^{-1/2} \sum_{t=1}^n ((TT')^{-1/2} - (T_k T_k')^{-1/2}) U_t^k$  converge in distribution to 0. Consequently we obtain that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} (TT')^{-1/2} U_t^k \xrightarrow[n \rightarrow \infty]{\mathbb{D}^{p+q+1}} B_{p+q+1}(r).$$

In order to conclude that (83) is true, it remains to observe that uniformly with respect to  $n$  it holds that

$$\tilde{Y}_n^k(r) := \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} (TT')^{-1/2} \tilde{Z}_t^k \xrightarrow[n \rightarrow \infty]{\mathbb{D}^{p+q+1}} 0, \quad (84)$$

where

$$\tilde{Z}_t^k = \left( -2 \left\{ \sum_{i=k+1}^{\infty} \dot{\lambda}_{i,1}(\theta_0) \epsilon_t \epsilon_{t-i}, \dots, \sum_{i=k+1}^{\infty} \dot{\lambda}_{i,p+q+1}(\theta_0) \epsilon_t \epsilon_{t-i} \right\} J^{-1'} \right)'.$$

By (64)

$$\sup_n \text{Var} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n \tilde{Z}_t^k \right) \xrightarrow{n \rightarrow \infty} 0$$

and since  $\lfloor nr \rfloor \leq n$ ,

$$\sup_{0 \leq r \leq 1} \sup_n \left\{ \left\| \tilde{Y}_n^k(r) \right\| \right\} \xrightarrow{n \rightarrow \infty} 0.$$

Thus (84) is true and the proof of (83) is achieved.

By (83) we deduce that

$$\frac{1}{\sqrt{n}} \left( \sum_{j=1}^{\lfloor nr \rfloor} (U_j - \bar{U}_n) \right) \xrightarrow[n \rightarrow \infty]{\mathbb{D}^{p+q+1}} (TT')^{1/2} (B_{p+q+1}(r) - rB_{p+q+1}(1)). \quad (85)$$

One remarks that the continuous mapping theorem on the Skorohod space yields

$$\begin{aligned} P_{p+q+1,n} &\xrightarrow[n \rightarrow \infty]{\text{in law}} (TT')^{1/2} \left[ \int_0^1 \{B_{p+q+1}(r) - rB_{p+q+1}(1)\} \{B_{p+q+1}(r) - rB_{p+q+1}(1)\}' dr \right] (TT')^{1/2} \\ &= (TT')^{1/2} V_{p+q+1} (TT')^{1/2}. \end{aligned}$$

Using (83), (85) and the continuous mapping theorem on the Skorohod space, one finally obtains

$$\begin{aligned} n(\hat{\theta}_n - \theta_0)' P_{p+q+1,n}^{-1} (\hat{\theta}_n - \theta_0) &\xrightarrow[n \rightarrow \infty]{\mathbb{D}^{p+q+1}} \left\{ (TT')^{1/2} B_{p+q+1}(1) \right\}' \left\{ (TT')^{1/2} V_{p+q+1} (TT')^{1/2} \right\}^{-1} \left\{ (TT')^{1/2} B_{p+q+1}(1) \right\} \\ &= B_{p+q+1}'(1) V_{p+q+1}^{-1} B_{p+q+1}(1) := \mathcal{U}_{p+q+1}. \end{aligned}$$

The proof of Theorem 5 is then complete.

### 6.7. Proof of Theorem 6

In view of (15) and (18), we write  $\hat{P}_{p+q+1,n} = P_{p+q+1,n} + Q_{p+q+1,n}$  where

$$\begin{aligned} Q_{p+q+1,n} &= (J(\theta_0)^{-1} - \hat{J}_n^{-1}) \frac{1}{n^2} \sum_{t=1}^n \left( \sum_{j=1}^t (H_j - \frac{1}{n} \sum_{k=1}^n H_k) \right) \left( \sum_{j=1}^t (H_j - \frac{1}{n} \sum_{k=1}^n H_k) \right)' \\ &\quad + \hat{J}_n^{-1} \frac{1}{n^2} \sum_{t=1}^n \left\{ \left( \sum_{j=1}^t (H_j - \frac{1}{n} \sum_{k=1}^n H_k) \right) \left( \sum_{j=1}^t (H_j - \frac{1}{n} \sum_{k=1}^n H_k) \right)' \right. \\ &\quad \left. - \left( \sum_{j=1}^t (\hat{H}_j - \frac{1}{n} \sum_{k=1}^n \hat{H}_k) \right) \left( \sum_{j=1}^t (\hat{H}_j - \frac{1}{n} \sum_{k=1}^n \hat{H}_k) \right)' \right\}. \end{aligned}$$

Using the same approach as in Lemma 8,  $\hat{J}_n$  converges almost surely to  $J$ . Thus we deduce that the first term in the right hand side of the above equation tends to zero in probability.

The second term is a sum composed of the following terms

$$q_{s,t}^{i,j,k,l} = \epsilon_s(\theta_0)\epsilon_t(\theta_0)\frac{\partial\epsilon_s(\theta_0)}{\partial\theta_i}\frac{\partial\epsilon_t(\theta_0)}{\partial\theta_j} - \tilde{\epsilon}_s(\hat{\theta}_n)\tilde{\epsilon}_t(\hat{\theta}_n)\frac{\partial\tilde{\epsilon}_s(\hat{\theta}_n)}{\partial\theta_k}\frac{\partial\tilde{\epsilon}_t(\hat{\theta}_n)}{\partial\theta_l}.$$

Using similar arguments done before (see for example the use of Taylor's expansion in Subsection 6.4, we have  $q_{s,t}^{i,j,k,l} = o_{\mathbb{P}}(1)$  as  $n$  goes to infinity and thus  $Q_{p+q+1,n} = o_{\mathbb{P}}(1)$ . So one may find a matrix  $Q_{p+q+1,n}^*$  that tends to the null matrix in probability and such that

$$\begin{aligned} n(\hat{\theta}_n - \theta_0)' \hat{P}_{p+q+1,n}^{-1}(\hat{\theta}_n - \theta_0) &= n(\hat{\theta}_n - \theta_0)' (P_{p+q+1,n} + Q_{p+q+1,n})^{-1}(\hat{\theta}_n - \theta_0) \\ &= n(\hat{\theta}_n - \theta_0)' P_{p+q+1,n}^{-1}(\hat{\theta}_n - \theta_0) \\ &\quad + n(\hat{\theta}_n - \theta_0)' Q_{p+q+1,n}^*(\hat{\theta}_n - \theta_0). \end{aligned}$$

Thanks to the arguments developed in the proof of Theorem 5,  $n(\hat{\theta}_n - \theta_0)' P_{p+q+1,n}^{-1}(\hat{\theta}_n - \theta_0)$  converges in distribution. So  $n(\hat{\theta}_n - \theta_0)' Q_{p+q+1,n}^*(\hat{\theta}_n - \theta_0)$  tends to zero in distribution, hence in probability. Then  $n(\hat{\theta}_n - \theta_0)' \hat{P}_{p+q+1,n}^{-1}(\hat{\theta}_n - \theta_0)$  and  $n(\hat{\theta}_n - \theta_0)' P_{p+q+1,n}^{-1}(\hat{\theta}_n - \theta_0)$  have the same limit in distribution and the result is proved.

## References

- [And71] T. W. ANDERSON – *The statistical analysis of time series*, John Wiley & Sons, Inc., New York-London-Sydney, 1971.
- [BCT96] R. T. BAILLIE, C.-F. CHUNG & M. A. TIESLAU – *Analysing inflation by the fractionally integrated ARFIMA-GARCH model*, *Journal of applied econometrics* **11** (1996), no. 1, p. 23–40.
- [Ber74] K. N. BERK – *Consistent autoregressive spectral estimates*, *Ann. Statist.* **2** (1974), p. 489–502, Collection of articles dedicated to Jerzy Neyman on his 80th birthday.
- [Ber95] J. BERAN – *Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models*, *J. Roy. Statist. Soc. Ser. B* **57** (1995), no. 4, p. 659–672.
- [BFGK13] J. BERAN, Y. FENG, S. GHOSH & R. KULIK – *Long-memory processes*, Springer, Heidelberg, 2013, Probabilistic properties and statistical methods.
- [BM09] Y. BOUBACAR MAÏNASSARA – *Estimation, validation et identification des modèles ARMA faibles multivariés*, PHD thesis of Université Lille 3, 2009.
- [BM12] Y. BOUBACAR MAÏNASSARA – *Selection of weak VARMA models by modified Akaike's information criteria*, *J. Time Series Anal.* **33** (2012), no. 1, p. 121–130.
- [BMCF12] Y. BOUBACAR MAINASSARA, M. CARBON & C. FRANCO – *Computing and estimating information matrices of weak ARMA models*, *Comput. Statist. Data Anal.* **56** (2012), no. 2, p. 345–361.
- [BMF11] Y. BOUBACAR MAINASSARA & C. FRANCO – *Estimating structural VARMA models with uncorrelated but non-independent error terms*, *J. Multivariate Anal.* **102** (2011), no. 3, p. 496–505.
- [BMK16] Y. BOUBACAR MAÏNASSARA & C. C. KOKONENDJI – *Modified Schwarz and Hannan-Quinn information criteria for weak VARMA models*, *Stat. Inference Stoch. Process.* **19** (2016), no. 2, p. 199–217.

- [BMS18] Y. BOUBACAR MAÏNASSARA & B. SAUSSEREAU – *Diagnostic checking in multivariate arma models with dependent errors using normalized residual autocorrelations*, *J. Amer. Statist. Assoc.* **113** (2018), no. 524, p. 1813–1827.
- [Bri81] D. R. BRILLINGER – *Time series: data analysis and theory*, vol. 36, Siam, 1981.
- [Chi88] S.-T. CHIU – *Weighted least squares estimators on the frequency domain for the parameters of a time series*, *Ann. Statist.* **16** (1988), no. 3, p. 1315–1326.
- [Dah89] R. DAHLHAUS – *Efficient parameter estimation for self-similar processes*, *Ann. Statist.* **17** (1989), no. 4, p. 1749–1766.
- [Dav94] J. DAVIDSON – *Stochastic limit theory*, Advanced Texts in Econometrics, The Clarendon Press, Oxford University Press, New York, 1994, An introduction for econometricians.
- [DGE93] Z. DING, C. W. GRANGER & R. F. ENGLE – *A long memory property of stock market returns and a new model*, *Journal of Empirical Finance* **1** (1993), p. 83–106.
- [DL89] P. DOUKHAN & J. LEÓN – *Cumulants for stationary mixing random sequences and applications to empirical spectral density*, *Probab. Math. Stat* **10** (1989), p. 11–26.
- [FRZ05] C. FRANCO, R. ROY & J.-M. ZAKOÏAN – *Diagnostic checking in ARMA models with uncorrelated errors*, *J. Amer. Statist. Assoc.* **100** (2005), no. 470, p. 532–544.
- [FT86] R. FOX & M. S. TAQQU – *Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series*, *Ann. Statist.* **14** (1986), no. 2, p. 517–532.
- [FY08] J. FAN & Q. YAO – *Nonlinear time series: nonparametric and parametric methods*, Springer Science & Business Media, 2008.
- [FZ98] C. FRANCO & J.-M. ZAKOÏAN – *Estimating linear representations of nonlinear processes*, *J. Statist. Plann. Inference* **68** (1998), no. 1, p. 145–165.
- [FZ05] — , *Recent results for linear time series models with non independent innovations*, in *Statistical modeling and analysis for complex data problems*, GERAD 25th Anniv. Ser., vol. 1, Springer, New York, 2005, p. 241–265.
- [FZ07] C. FRANCO & J.-M. ZAKOÏAN – *HAC estimation and strong linearity testing in weak ARMA models*, *J. Multivariate Anal.* **98** (2007), no. 1, p. 114–144.
- [FZ10] C. FRANCO & J.-M. ZAKOÏAN – *GARCH models: Structure, statistical inference and financial applications*, Wiley, 2010.
- [GJ80] C. W. J. GRANGER & R. JOYEUX – *An introduction to long-memory time series models and fractional differencing*, *J. Time Ser. Anal.* **1** (1980), no. 1, p. 15–29.
- [GS90] L. GIRAITIS & D. SURGAILIS – *A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate*, *Probab. Theory Related Fields* **86** (1990), no. 1, p. 87–104.
- [Her84] N. HERRNDORF – *A functional central limit theorem for weakly dependent sequences of random variables*, *Ann. Probab.* **12** (1984), no. 1, p. 141–153.
- [HK98] M. HAUSER & R. KUNST – *Fractionally integrated models with arch errors: with an application to the swiss 1-month euromarket interest rate*, *Review of Quantitative Finance and Accounting* **10** (1998), no. 1, p. 95–113.
- [Hos81] J. R. M. HOSKING – *Fractional differencing*, *Biometrika* **68** (1981), no. 1, p. 165–176.
- [HTSC99] M. HALLIN, M. TANIGUCHI, A. SERROUKH & K. CHOY – *Local asymptotic normality for regression models with long-memory disturbance*, *Ann. Statist.* **27** (1999), no. 6, p. 2054–2080.
- [Kee87] D. M. KEENAN – *Limiting behavior of functionals of higher-order sample cumulant spectra*, *Ann. Statist.* **15** (1987), no. 1, p. 134–151.
- [KL06] C.-M. KUAN & W.-M. LEE – *Robust M tests without consistent estimation of the*

- asymptotic covariance matrix, *J. Amer. Statist. Assoc.* **101** (2006), no. 475, p. 1264–1275.
- [LL97] S. LING & W. K. LI – *On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity*, *J. Amer. Statist. Assoc.* **92** (1997), no. 439, p. 1184–1194.
- [Lob01] I. N. LOBATO – *Testing that a dependent process is uncorrelated*, *J. Amer. Statist. Assoc.* **96** (2001), no. 455, p. 1066–1076.
- [Lüt05] H. LÜTKEPOHL – *New introduction to multiple time series analysis*, Springer-Verlag, Berlin, 2005.
- [MN99] J. R. MAGNUS & H. NEUDECKER – *Matrix differential calculus with applications in statistics and econometrics*, Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester, 1999, Revised reprint of the 1988 original.
- [Pal07] W. PALMA – *Long-memory time series*, Wiley Series in Probability and Statistics, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2007, Theory and methods.
- [Sha10a] X. SHAO – *Corrigendum: A self-normalized approach to confidence interval construction in time series*, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** (2010), no. 5, p. 695–696.
- [Sha10b] — , *Nonstationarity-extended Whittle estimation*, *Econometric Theory* **26** (2010), no. 4, p. 1060–1087.
- [Sha10c] — , *A self-normalized approach to confidence interval construction in time series*, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** (2010), no. 3, p. 343–366.
- [Sha11] — , *Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models*, *Econometric Theory* **27** (2011), no. 2, p. 312–343.
- [Sha12] — , *Parametric inference in stationary time series models with dependent errors*, *Scand. J. Stat.* **39** (2012), no. 4, p. 772–783.
- [Sha15] — , *Self-normalization for time series: a review of recent developments*, *J. Amer. Statist. Assoc.* **110** (2015), no. 512, p. 1797–1817.
- [Tan82] M. TANIGUCHI – *On estimation of the integrals of the fourth order cumulant spectral density*, *Biometrika* **69** (1982), no. 1, p. 117–122.
- [Ton90] H. TONG – *Non-linear time series: a dynamical system approach*, Oxford University Press, 1990.
- [TT97] M. S. TAQQU & V. TEVEROVSKY – *Robustness of Whittle-type estimators for time series with long-range dependence*, *Comm. Statist. Stochastic Models* **13** (1997), no. 4, p. 723–757, Heavy tails and highly volatile phenomena.
- [Whi53] P. WHITTLE – *Estimation and information in stationary time series*, *Ark. Mat.* **2** (1953), p. 423–434.
- [ZL15] K. ZHU & W. K. LI – *A bootstrapped spectral test for adequacy in weak ARMA models*, *J. Econometrics* **187** (2015), no. 1, p. 113–130.



## 7. Figures and tables

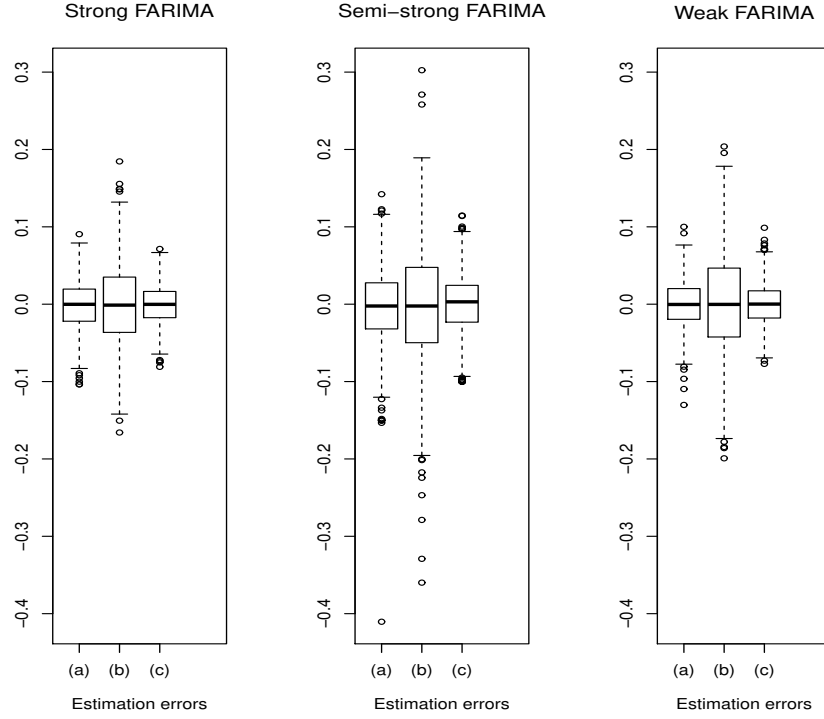


FIG 1. LSE of  $N = 1,000$  independent simulations of the FARIMA(1,  $d$ , 1) model (19) with size  $n = 2,000$  and unknown parameter  $\theta_0 = (a, b, d) = (-0.7, -0.2, 0.4)$ , when the noise is strong (left panel), when the noise is semi-strong (20) (middle panel) and when the noise is weak of the form (21) (right panel). Points (a)-(c), in the box-plots, display the distribution of the estimation error  $\hat{\theta}_n(i) - \theta_0(i)$  for  $i = 1, 2, 3$ .

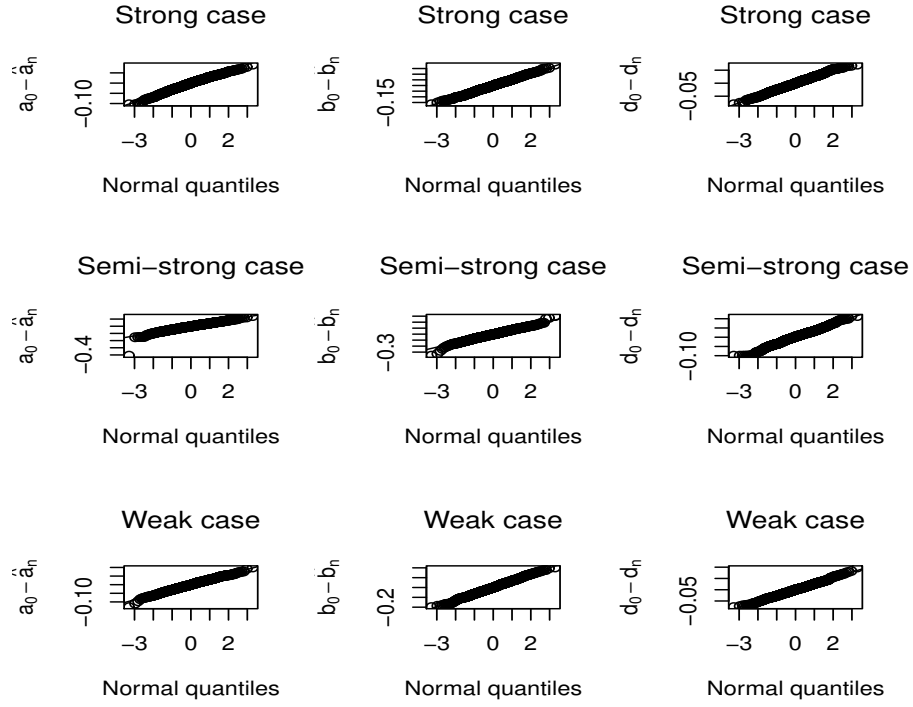


FIG 2. LSE of  $N = 1,000$  independent simulations of the FARIMA(1,  $d$ , 1) model (19) with size  $n = 2,000$  and unknown parameter  $\theta_0 = (a, b, d) = (-0.7, -0.2, 0.4)$ . The top panels present respectively, from left to right, the Q-Q plot of the estimates  $\hat{a}_n$ ,  $\hat{b}_n$  and  $\hat{d}_n$  of  $a$ ,  $b$  and  $d$  in the strong case. Similarly the middle and the bottom panels present respectively, from left to right, the Q-Q plot of the estimates  $\hat{a}_n$ ,  $\hat{b}_n$  and  $\hat{d}_n$  of  $a$ ,  $b$  and  $d$  in the semi-strong and weak cases.

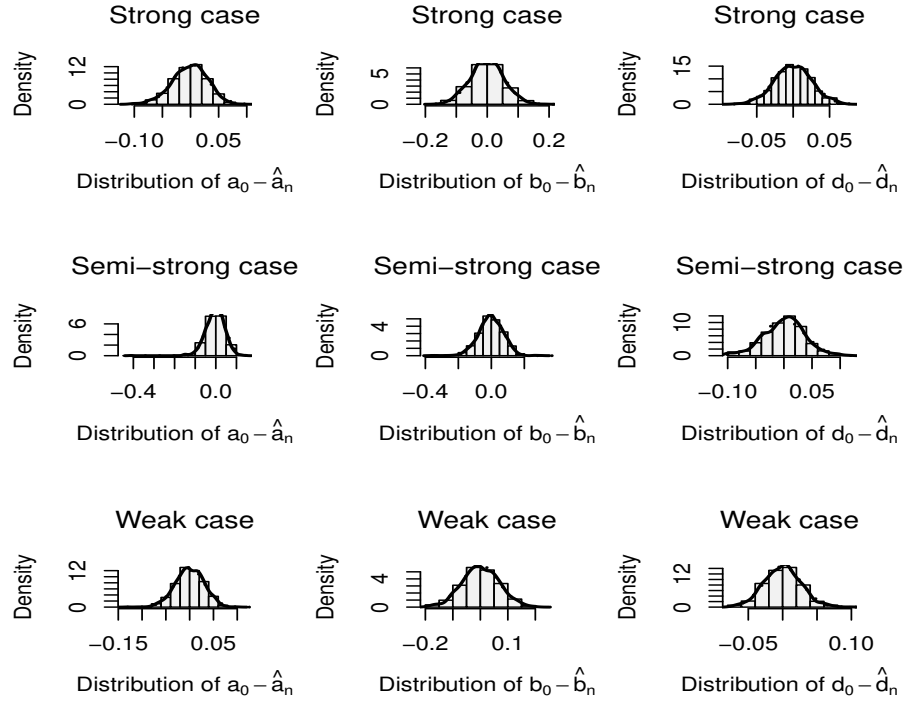


FIG 3. LSE of  $N = 1,000$  independent simulations of the FARIMA(1,  $d$ , 1) model (19) with size  $n = 2,000$  and unknown parameter  $\theta_0 = (a, b, d) = (-0.7, -0.2, 0.4)$ . The top panels present respectively, from left to right, the distribution of the estimates  $\hat{a}_n$ ,  $\hat{b}_n$  and  $\hat{d}_n$  of  $a$ ,  $b$  and  $d$  in the strong case. Similarly the middle and the bottom panels present respectively, from left to right, the distribution of the estimates  $\hat{a}_n$ ,  $\hat{b}_n$  and  $\hat{d}_n$  of  $a$ ,  $b$  and  $d$  in the semi-strong and weak cases. The kernel density estimate is displayed in full line, and the centered Gaussian density with the same variance is plotted in dotted line.

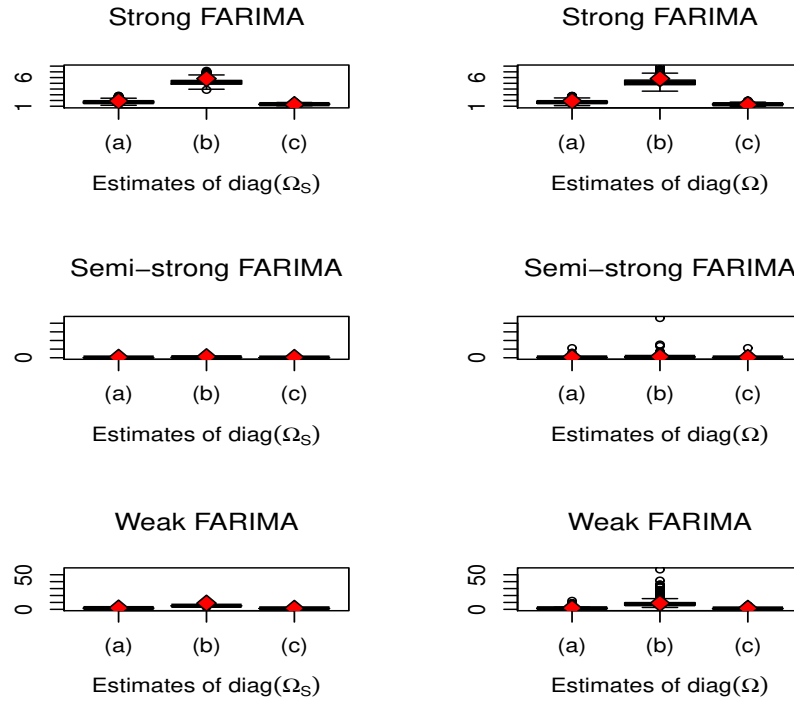


FIG 4. Comparison of standard and modified estimates of the asymptotic variance  $\Omega$  of the LSE, on the simulated models presented in Figure 1. The diamond symbols represent the mean, over  $N = 1,000$  replications, of the standardized errors  $n(\hat{a}_n + 0.7)^2$  for (a) (1.90 in the strong case and 4.32 (resp. 1.80) in the semi-strong case (resp. in the weak case)),  $n(\hat{b}_n + 0.2)^2$  for (b) (5.81 in the strong case and 11.33 (resp. 8.88) in the semi-strong case (resp. in the weak case)) and  $n(\hat{d}_n - 0.4)^2$  for (c) (1.28 in the strong case and 2.65 (resp. 1.40) in the semi-strong case (resp. in the weak case)).

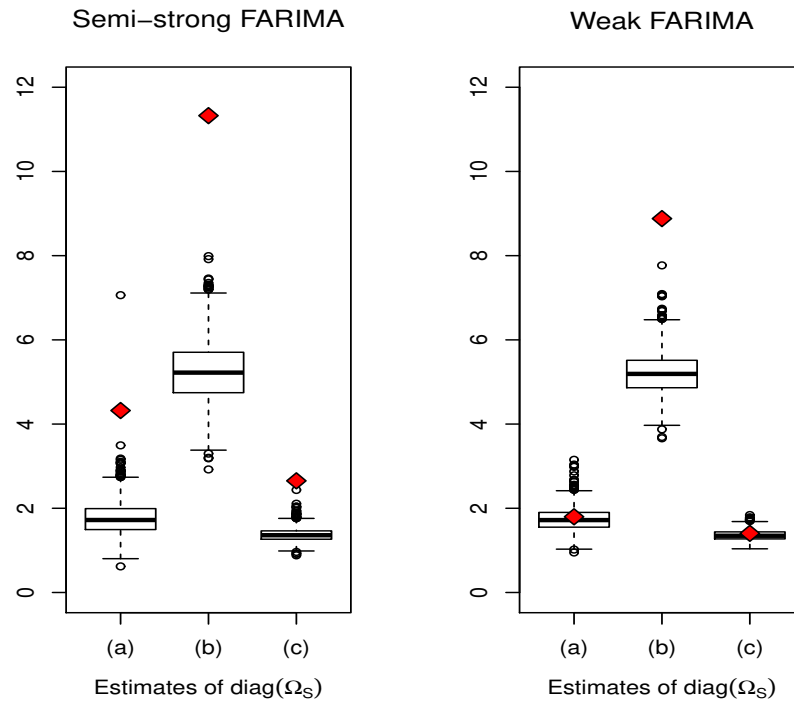


FIG 5. A zoom of the left-middle and left-bottom panels of Figure 4

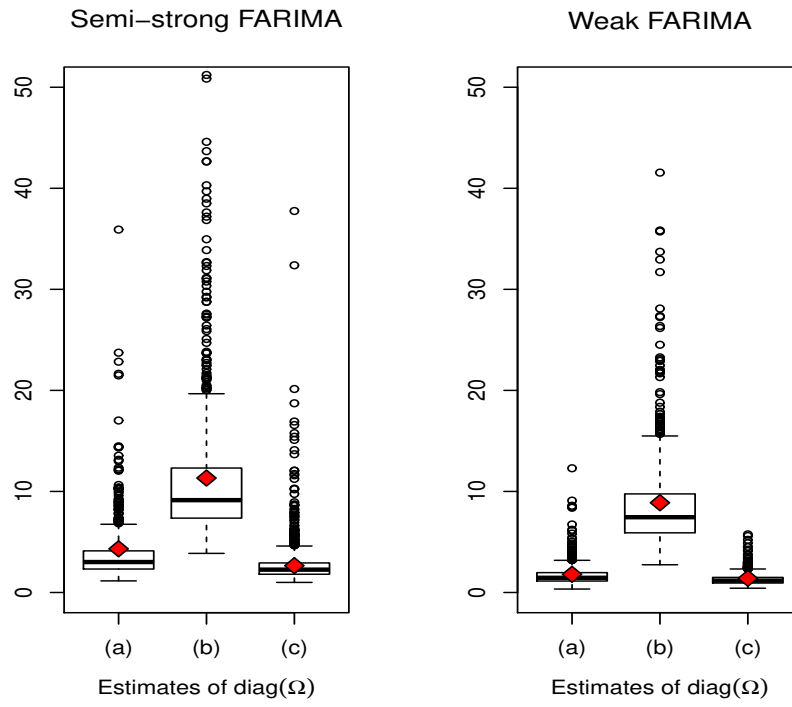


FIG 6. A zoom of the right-middle and right-bottom panels of Figure 4

TABLE 1

Empirical size of standard and modified confidence interval: relative frequencies (in %) of rejection. Modified SN stands for the self-normalized approach. In Modified we use the sandwich estimator of the asymptotic variance  $\Omega$  of the LSE while in Standard we use  $\hat{\Omega}_S$ . The number of replications is  $N = 1000$ .

Model	Length $n$	Level	Standard			Modified			Modified SN		
			$\hat{a}_n$	$\hat{b}_n$	$\hat{d}_n$	$\hat{a}_n$	$\hat{b}_n$	$\hat{d}_n$	$\hat{a}_n$	$\hat{b}_n$	$\hat{d}_n$
Strong FARIMA	$n = 200$	$\alpha = 1\%$	<b>2.8</b>	<b>2.7</b>	<b>2.1</b>	<b>3.7</b>	<b>3.1</b>	<b>2.5</b>	<b>2.5</b>	<b>2.5</b>	<b>2.0</b>
		$\alpha = 5\%$	<b>7.1</b>	<b>7.3</b>	5.2	<b>8.2</b>	<b>8.0</b>	5.4	<b>8.4</b>	<b>6.9</b>	5.6
		$\alpha = 10\%$	11.8	11.2	8.3	<b>12.8</b>	<b>12.4</b>	9.5	<b>14.5</b>	11.5	10.6
Strong FARIMA	$n = 2,000$	$\alpha = 1\%$	1.1	1.6	0.7	1.3	1.6	1.0	1.6	1.0	0.8
		$\alpha = 5\%$	5.8	<b>6.9</b>	5.1	6.1	<b>6.8</b>	5.3	5.6	6.4	3.8
		$\alpha = 10\%$	10.9	<b>13.0</b>	9.5	11.4	<b>12.8</b>	9.5	10.3	<b>12.4</b>	9.0
Strong FARIMA	$n = 5,000$	$\alpha = 1\%$	1.3	1.2	0.7	1.2	1.2	0.8	0.8	1.3	1.2
		$\alpha = 5\%$	5.3	4.9	5.2	5.7	5.1	5.4	5.1	4.9	5.3
		$\alpha = 10\%$	10.6	10.3	11.4	10.7	10.2	11.7	11.8	10.8	11.4
Semi-strong FARIMA	$n = 200$	$\alpha = 1\%$	<b>5.3</b>	<b>3.7</b>	<b>2.3</b>	<b>4.8</b>	<b>3.4</b>	<b>3.2</b>	<b>4.0</b>	<b>2.8</b>	1.4
		$\alpha = 5\%$	<b>11.2</b>	<b>9.5</b>	6.1	<b>10.7</b>	<b>9.1</b>	5.9	<b>11.1</b>	<b>8.5</b>	5.7
		$\alpha = 10\%$	<b>16.8</b>	<b>14.5</b>	8.6	<b>16.7</b>	<b>14.7</b>	10.0	<b>17.1</b>	<b>13.9</b>	10.9
Semi-strong FARIMA	$n = 2,000$	$\alpha = 1\%$	<b>6.8</b>	<b>7.7</b>	<b>7.8</b>	1.7	0.9	1.4	<b>2.3</b>	<b>2.0</b>	0.8
		$\alpha = 5\%$	<b>19.5</b>	<b>17.7</b>	<b>14.9</b>	<b>6.5</b>	5.8	5.5	<b>8.1</b>	<b>6.8</b>	<b>6.5</b>
		$\alpha = 10\%$	<b>26.5</b>	<b>26.7</b>	<b>21.5</b>	<b>13.5</b>	11.0	9.9	<b>14.6</b>	<b>12.8</b>	<b>12.5</b>
Semi-strong FARIMA	$n = 5,000$	$\alpha = 1\%$	<b>11.2</b>	<b>9.8</b>	<b>9.4</b>	1.6	1.5	1.1	1.6	1.3	1.2
		$\alpha = 5\%$	<b>20.8</b>	<b>20.2</b>	<b>20.9</b>	6.4	5.7	5.3	5.7	6.2	<b>7.2</b>
		$\alpha = 10\%$	<b>28.2</b>	<b>28.4</b>	<b>28.4</b>	<b>12.2</b>	9.8	11.3	<b>12.0</b>	<b>13.0</b>	<b>13.9</b>
Weak FARIMA	$n = 200$	$\alpha = 1\%$	<b>2.6</b>	<b>4.4</b>	1.2	<b>6.2</b>	<b>6.9</b>	<b>4.3</b>	<b>4.1</b>	<b>4.2</b>	<b>2.6</b>
		$\alpha = 5\%$	<b>6.6</b>	<b>11.3</b>	4.3	<b>13.8</b>	<b>14.6</b>	<b>10.2</b>	<b>12.0</b>	<b>10.5</b>	<b>8.9</b>
		$\alpha = 10\%$	10.9	<b>18.8</b>	<b>7.1</b>	<b>20.3</b>	<b>21.9</b>	<b>16.3</b>	<b>17.7</b>	17.4	<b>15.4</b>
Weak FARIMA	$n = 2,000$	$\alpha = 1\%$	1.1	<b>5.3</b>	1.3	1.5	1.2	1.6	1.2	1.1	0.9
		$\alpha = 5\%$	5.4	<b>13.4</b>	5.8	<b>7.0</b>	<b>6.8</b>	5.5	5.7	<b>6.5</b>	6.4
		$\alpha = 10\%$	11.4	<b>21.2</b>	9.6	<b>12.8</b>	<b>12.0</b>	11.2	11.3	<b>11.9</b>	<b>12.2</b>
Weak FARIMA	$n = 5,000$	$\alpha = 1\%$	1.3	<b>4.6</b>	1.7	1.2	1.3	1.2	1.3	1.4	0.9
		$\alpha = 5\%$	6.3	<b>14.4</b>	6.0	<b>6.7</b>	6.3	5.9	6.2	6.2	5.0
		$\alpha = 10\%$	11.5	<b>22.3</b>	11.6	<b>12.1</b>	<b>12.3</b>	10.8	10.6	10.9	10.0

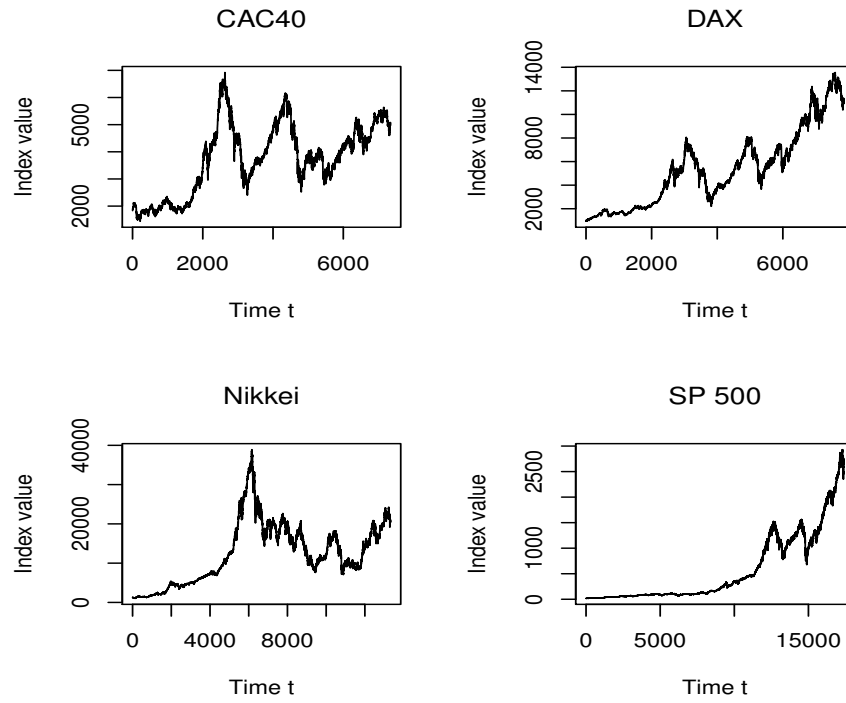


FIG 7. Closing prices of the four stock market indices from the starting date of each index to February 14, 2019.



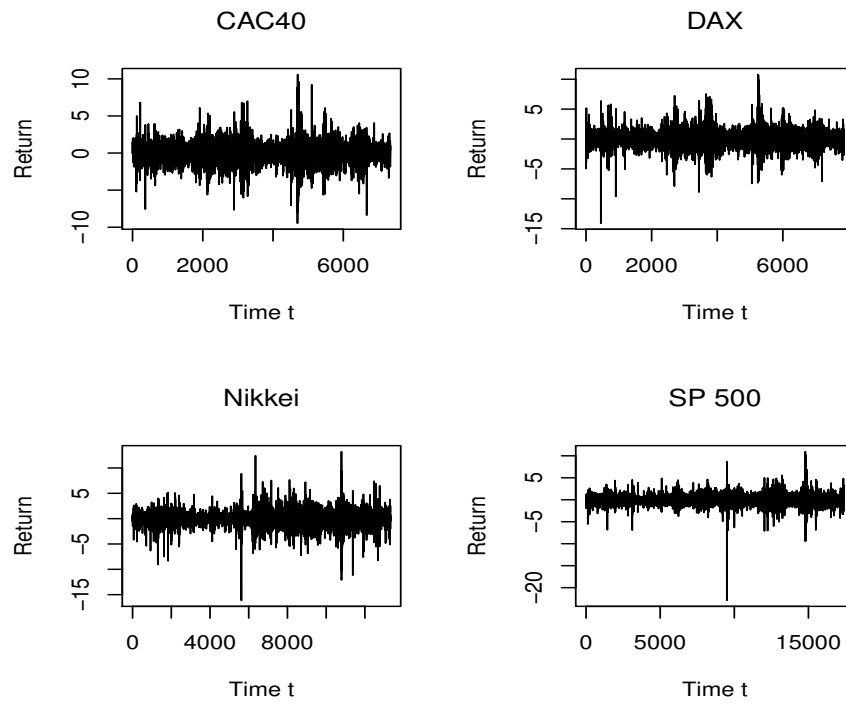


FIG 8. Returns of the four stock market indices from the starting date of each index to February 14, 2019.

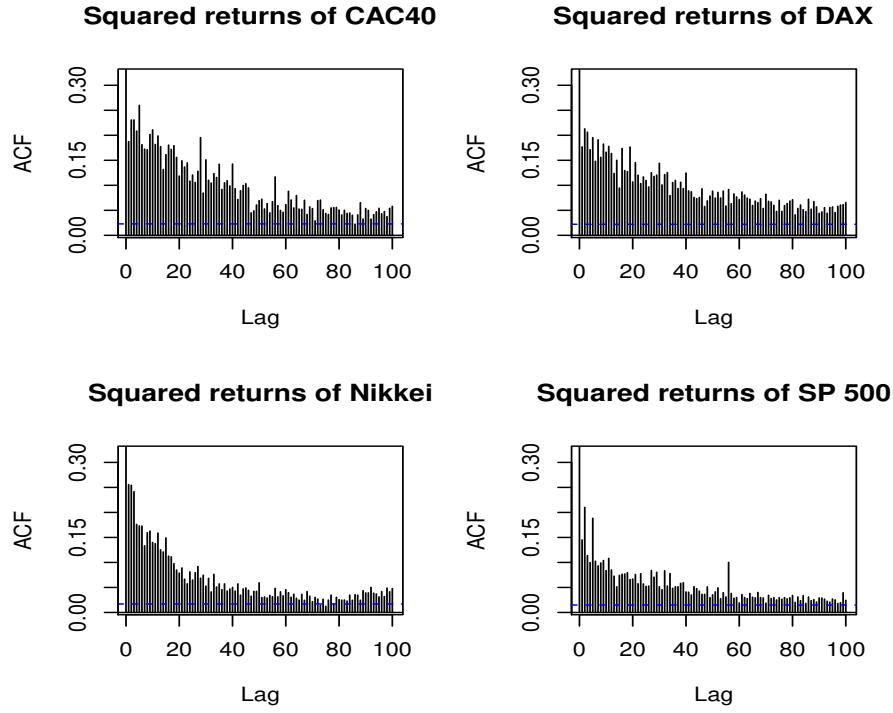


FIG 9. Sample autocorrelations of squared returns of the four stock market indices.

TABLE 2

Fitting a  $FARIMA(1, d, 1)$  model to the squares of the 4 daily returns considered. The corresponding  $p$ -values are reported in parentheses. The last column presents the estimated residual variance.

Series	Length $n$	$\hat{\theta}_n$			$\text{Var}(\epsilon_t)$ $\hat{\sigma}_\epsilon^2$
		$\hat{a}_n$	$\hat{b}_n$	$\hat{d}_n$	
CAC	$n = 7,341$	0.1199 (0.1524)	0.5296 (0.0000)	0.4506 (0.0000)	$19.6244 \times 10^{-8}$
DAX	$n = 7,860$	0.1598 (0.1819)	0.4926 (0.0000)	0.3894 (0.0000)	$25.9383 \times 10^{-8}$
Nikkei	$n = 13,318$	-0.0217 (0.9528)	0.1579 (0.6050)	0.3217 (0.0000)	$25.6844 \times 10^{-8}$
S&P 500	$n = 17,390$	-0.3371 (0.0023)	-0.1795 (0.0227)	0.2338 (0.0000)	$22.9076 \times 10^{-8}$

TABLE 3

Modified confidence interval at the asymptotic level  $\alpha = 5\%$  for the parameters estimated in Table 2.

Modified SN stands for the self-normalized approach while Modified corresponds to the confidence interval obtained by using the sandwich estimator of the asymptotic variance  $\Omega$  of the LSE.

Series	Modified			Modified SN		
	$\hat{a}_n$	$\hat{b}_n$	$\hat{d}_n$	$\hat{a}_n$	$\hat{b}_n$	$\hat{d}_n$
CAC	$[-0.044, 0.284]$	$[0.319, 0.740]$	$[0.251, 0.651]$	$[0.049, 0.193]$	$[0.432, 0.627]$	$[0.358, 0.543]$
DAX	$[-0.075, 0.394]$	$[0.272, 0.714]$	$[0.263, 0.516]$	$[-0.219, 0.537]$	$[0.422, 0.563]$	$[0.158, 0.621]$
Nikkei	$[-0.741, 0.698]$	$[-0.440, 0.756]$	$[0.206, 0.437]$	$[-0.823, 0.779]$	$[-0.407, 0.717]$	$[0.156, 0.488]$
S&P 500	$[-0.554, -0.121]$	$[-0.334, -0.025]$	$[0.162, 0.306]$	$[-0.430, -0.244]$	$[-0.297, -0.062]$	$[0.101, 0.366]$

## Contents

1	Introduction . . . . .	1
2	Least squares estimation . . . . .	4
2.1	Notations and assumptions . . . . .	4
2.2	Asymptotic properties . . . . .	6
3	Estimating the asymptotic variance matrix . . . . .	7
3.1	Estimation of the asymptotic matrix $I(\theta_0)$ . . . . .	7
3.2	A self-normalized approach to confidence interval construction in weak FARIMA models . . . . .	9
4	Numerical illustrations . . . . .	11
4.1	Simulation studies and empirical sizes for confidence intervals . . . . .	11
4.2	Application to real data . . . . .	12
5	Conclusion . . . . .	13
6	Proofs . . . . .	13
6.1	Preliminary results . . . . .	13
6.2	Proof of Theorem 1 . . . . .	17
6.3	Proof of Theorem 2 . . . . .	18
6.4	Proof of the convergence of the variance matrix estimator . . . . .	33
6.5	Invertibility of the normalization matrix $P_{p+q+1,n}$ . . . . .	45
6.6	Proof of Theorem 5 . . . . .	46
6.7	Proof of Theorem 6 . . . . .	48
	References . . . . .	49
7	Figures and tables . . . . .	52